

THIS WEEK

EDITORIALS

EUROPE The continent broadens its approach to science **p.288**



WORLD VIEW Deepwater Horizon two years on — what now? **p.289**

CORAL Flexible reefs could cope with climate change **p.290**

Justice for all

The US government must not wriggle out of paying compensation to the victims of horrific experiments in Guatemala in the 1940s.

The US Department of Justice is trying to persuade a court in Washington DC to throw out a lawsuit filed last year by the survivors of shameful 1940s experiments in Guatemala that were funded by the US government. More than 1,300 Guatemalans, almost all of them poor or otherwise vulnerable, were intentionally exposed to the diseases syphilis, gonorrhoea or chancroid without their consent. The experiments also took blood and cerebrospinal fluid from 5,128 adults and children, again with no consent. Of those infected, unknown numbers contracted the diseases and died; others suffered for decades, in some cases infecting their spouses or their children (see *Nature* **482**, 148–152; 2012).

US President Barack Obama, Secretary of State Hillary Clinton and Secretary of Health and Human Services Kathleen Sebelius were quick to issue public apologies when news of the heinous experiments on prisoners, prostitutes, orphans, soldiers and patients with leprosy or mental illness broke in 2010. Obama also asked his bioethics commission to investigate. It issued a damning, 200-page report last September.

But when lawyers for the victims asked US Attorney General Eric Holder to set up a claims process that would meaningfully compensate the survivors, they hit a wall of silence.

So in March 2011, they sued senior officials at agencies including the National Institutes of Health in Bethesda, Maryland, and the Centers for Disease Control and Prevention in Atlanta, Georgia, asking for compensatory and punitive damages. These agencies and others funded and oversaw the Guatemalan experiments in the name of science 65 years ago.

In a court filing last week, the US government reiterated a stance that it first took in January: that the court should throw the case out because government officials are shielded from lawsuits that arise from actions taken in the course of performing their jobs, and because current officials cannot be held responsible for the acts of their predecessors decades ago.

The United States may be in the right, legally. It can be argued that no one would work for the government if they could be sued every time their actions, never mind those of their predecessors, had a negative impact on a member of the public. But the government is in the wrong, morally, in refusing to find a way to compensate the survivors. These people and their families have borne decades of pain, misery and poverty on the heels of experiments called “reprehensible” in a joint statement by Clinton and Sebelius, and for which Obama, in a personal phone call with the Guatemalan president, expressed his “deep regret”.

These apologies will ring hollow until officials establish a compensation fund that is outside the court system. (The courts could take months or years to reach a conclusion in this case, and time is of the essence: already, one of the elderly plaintiffs has died since the lawsuit was filed.)

There is plenty of precedent for reparations for unethical medical experimentation. The United States paid US\$37,500 and lifetime

health benefits to each living survivor of the Tuskegee syphilis experiments, which involved 399 poor African American tenant farmers in Alabama. Those US-funded experiments left the syphilis-infected farmers untreated between 1932 and 1972.

In 1996, the US government paid \$4.8 million in compensation for injecting 12 people, most of them hospital patients being treated for unrelated illnesses, with plutonium and uranium in the mid-1940s. Those compensated included one survivor and 11 family members of deceased victims. And earlier this year, a task force in North Carolina recommended that victims of that state’s forced-sterilization programme, which ended in 1974, each receive \$50,000.

“There is a compelling moral case that the US government owes reparations to the Guatemalans.”

There is a compelling moral case that the US government — a continuous institution since the time of the horrific Guatemalan experiments — owes reparations to the

Guatemalans affected. Senior officials clearly recognize this: the day after the Department of Justice first asked the court to throw out the case, the Department of Health and Human Services announced that it would spend \$775,000 to prevent and treat sexually transmitted diseases in Guatemala. It pledged another \$1 million to evaluate the effectiveness of changes being made to the way human subjects are protected in medical experiments.

These measures are necessary, but not sufficient. The administration must find a way to put money in the pockets of those who paid the highest price for these evil acts — and quickly. If it cannot, or will not, Congress should step into the breach. Justice demands no less. ■

No shame

The handling of results suggesting faster-than-light neutrinos was a model of fitting behaviour.

If the public learned one thing about physics last year, it was that a particle had been found that might travel faster than the speed of light. Most people were probably vague about what the particle was, but they seemed to grasp the significance. The Universe’s speed limit was in doubt, and anything might be possible. The result, announced by scientists at the OPERA neutrino experiment in Gran Sasso, Italy, may have been wrong, but the message conveyed about science was not. Late last month, following a vote of no-confidence in their leadership, OPERA’s two top scientists resigned. Yet both men, along with the rest of the collaboration, can hold their heads high.

The vote and the resignations have not been officially blamed on the media circus over the faster-than-light neutrinos, which OPERA brought to public attention and was then forced to admit did not exist after all. But that is how they will be interpreted. And if the vote is portrayed as a referendum on how OPERA handled the situation, then scientists everywhere should think carefully about how they would have voted. What kind of science do they want?

The neutrino story is familiar to most researchers now, but here are the highlights. OPERA was measuring a beam of neutrinos coming from CERN, Europe's high-energy physics lab near Geneva, Switzerland. Contrary to everything taught in modern physics, the neutrinos seemed to be arriving 60 nanoseconds faster than light speed. A small sub-team of researchers responsible for the measurement spent months systematically checking OPERA's detector and could find no reason for the discrepancy.

When the smaller group shared their result with the full OPERA collaboration, it leaked to the Italian press. Faced with growing interest, OPERA's leaders — Antonio Ereditato and Dario Auterio, the duo who have now resigned — decided to go public with a seminar.

Physicists saw plenty of reasons to doubt OPERA's extraordinary claim, and Ereditato and Auterio did not disagree. Even as they presented the result, they invited their colleagues to comment, and encouraged others to try to reproduce their results. Within months, CERN had sent a new beam of neutrinos to Italy and a second experiment found neutrinos travelling at the expected speed. After a great deal of searching, members of the OPERA group eventually traced the discrepancy to a cable that was not fully screwed in.

Scientists both inside OPERA and out have since fretted about what such a high-profile misstep might mean for funding, reputation and the public's perception of science. In fact, OPERA's handling of the incident, at least publicly, was a model for how scientists should behave. Ereditato and Auterio acted responsibly when speaking publicly by sticking close to their data and avoiding over-interpretation. They shared their work with their competitors, and did their best to

quickly address outside criticism. In the end, it was OPERA's internal checks that found the loose cable. When the error was discovered, physicists on the team wasted no time in publicly announcing the problem, along with others they had exposed during their review.

Broadly speaking, the media and the public seemed to grasp that this is the way science is supposed to work. Some Italian journalists aside, the press responsibly reported the initial result as simultaneously

"The message here is that scientists are not afraid to question the big ideas."

incredible and very possibly wrong. The public enjoyed the opportunity to question the world around them and learned a little physics in the process. Media coverage generally sided with the researchers for admitting they were wrong, and no one has called for funding to be cut.

Science can fall victim to human frailties.

One researcher hoards her samples out of fear of competition; another doggedly promotes his hypothesis long after the data have falsified it; negative results are hidden because of competing financial interests. And the most frequent sin of all: questionable results go unchecked because it is in nobody's interest to check them.

The OPERA collaboration is not exempt from the human condition. Some collaborators believe that publication was rushed out of a desire to beat the competition. But OPERA nevertheless conducted itself openly and properly.

The no-confidence vote and resignations are a matter for the collaboration's internal processes, and have no bearing on the quality of the collaboration's science. But beyond OPERA itself, scientists should celebrate the way in which the results were disseminated and the findings ultimately refuted. The process was open and deliberate, and it led to the correct scientific result. In an era in which politics, business and celebrity fixate on spin, control and staying 'on message', OPERA's rise and fall make science stand apart. The message here is that scientists are not afraid to question the big ideas. They are not afraid to open themselves to public scrutiny. And they should not be afraid to be wrong. ■

All together now

The financial crisis brings Europe unexpected opportunities for international collaboration.

Big pieces of scientific kit aren't getting any cheaper. Neutron sources, synchrotrons, telescopes and particle accelerators all require highly trained staff and lots of consumables. They need to be upgraded periodically for research to march forward, and the next generation must be even bigger and more ambitious.

In the current fiscal environment, just keeping the lights on can be a struggle, but Europe's facilities are finding a way. As we report on page 295, they have begun inviting non-European countries to join them. The decision is driven by short-term need, but the continent's organizations may also be positioning themselves for an era of scientific supremacy over rivals such as the United States and Japan.

Europe understands the need for cooperation. Multinational labs bloomed throughout the second half of the twentieth century, a sign of both Europe's post-war unity and the fact that no single nation could afford the types of facility that were being built elsewhere in the world. Fast-forward 50 years, and bodies such as CERN, Europe's high-energy physics lab near Geneva, Switzerland, and the European Southern Observatory (ESO), based in Garching, Germany, are internationally recognized. They have built instruments that are among the best in the world, and they are training a generation of European scientists.

This year sees non-European members entering the fold for the first time. Last month, Israel became the first state from outside Europe to

be represented on CERN's governing council, and ESO will soon admit Brazil. Other bodies, including some facilities now under construction, are partnering with countries such as India. Even national facilities are joining in: after years of cooperation, Italy's Elettra synchrotron light source near Trieste has partnered with India to build two beamlines.

This expansion attracts cash in the short term, but it also lays the groundwork for a more fruitful future. Organizations such as ESO and CERN will gain the best researchers from nations around the world, raising their status and the quality of their research. They will also strengthen the scientific base in their partner countries, increasing the chances that these nations will be politically willing and scientifically able to help with the next big project. In the United States and Japan, by contrast, large national facilities are run by powerful research agencies that are difficult to approach and poorly equipped to engage in major financial and material cooperation at home and abroad. They will lose out in future: the facilities of tomorrow will be so big and costly that only those who can cooperate will win.

Two examples show the strength of Europe's hand. The first is ITER, a multinational fusion experiment that is being built in southern France. ITER's ballooning budget is giving Europe headaches in the short term, but its long-term benefits to the continent's scientific and technical expertise are undeniable. The second project, the Square Kilometre Array radio telescope, will be not be located in Europe for scientific reasons, but European scientists have a central role in choosing where it will be built, and will figure prominently in its operation.

There are risks, of course: chiefly that Europe will cede some control of its own infrastructure to its new partners. But done properly, the shared benefits can keep Europe on the top of the world for decades. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunqv

C. NEMAZIE



Deep-water drilling remains a risky business

Donald Boesch argues that the lessons learned from the Deepwater Horizon oil spill should be enshrined in legislation before they fade from memory.

Two years after the blowout of the BP oil well drilled by the Deepwater Horizon rig in the Gulf of Mexico, the United States is largely failing to act on the lessons learned from that experience to ensure that deep-water drilling and production is safe and environmentally compatible. In particular, the US Congress has not enacted any legislation to improve safety and protect the environment. Meanwhile, high oil prices are stimulating the expansion of drilling into ever deeper waters in the Gulf of Mexico, as well as off Brazil, Africa and Europe. Drilling is also proceeding in shallower, but ice-prone, regions of the Arctic, including the Beaufort and Chukchi seas off Alaska.

I am one of two scientists who served on the US president's commission that produced the report *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling* — the other was Cherry Murray of Harvard University. The commission concluded that the root causes of the blowout and explosion were deficiencies in regulatory oversight and multiple poor decisions made in the absence of a comprehensive risk-management system. Other investigations into the disaster essentially agreed.

We were impressed by the technologies developed to produce hydrocarbons from ever deeper, more highly pressured formations, but surprised by the lack of sophistication in techniques for detecting and controlling risk, containing the flow of hydrocarbons, collecting spilled oil and protecting vulnerable resources. For example, 'down-hole' events — those taking place deep below the seabed — are often inferred from indirect measurements of pressure and volume rather than measured with state-of-the-art *in situ* sensors of the type used in geophysical research and other industries.

Cement formulation, testing and placement — major factors in the blowout — seem to be more of an art than a science. Cementing is also central to debates on the increased recovery of hydrocarbons by hydrofracturing, because it is critical both to limiting fugitive emissions of methane and to preventing contamination of shallower aquifers.

The 2010 accident showed that no operating company in the world had the capacity to rapidly contain a deep-water blowout. It took months of seat-of-the-pants engineering to build and deploy a capping stack that provided effective containment. Confusion reigned over the fate of the oil and gas released 1,500 metres below the surface, largely because of a lack of understanding of the operating environment, including the direction and speed of water currents, and the behaviour of hydrocarbons released at depth.

There have been some positive responses to the Deepwater Horizon experience. The US Department of the Interior (DOI) temporarily suspended deep-water drilling in the Gulf until

new safeguards, including a demonstrated capacity to contain blowouts, were put into place. The DOI also reformed its management and oversight of offshore oil and gas development, separating safety regulation from developmental decision-making, and has established a new safety and environmental management system.

The oil industry, in addition to developing the needed containment capacity, has improved its safety processes. The American Petroleum Institute, the industry body, has created a Center for Offshore Safety and named Charlie Williams, a seasoned and respected scientist from Royal Dutch Shell, to head it. We shall see whether this new centre can develop the planned third-party audit process and if the industry, working with the DOI, will advance cutting-edge research and development (R&D) of safety technology.

BP has funded an independently managed Gulf Research Initiative to support longer-term research on the impacts of the Deepwater Horizon oil spill. Through this, and the natural resources damage assessment being conducted by state and federal agencies, we will learn much about the fate and the effects of the hydrocarbons released during the blowout. And a task force established by President Barack Obama has developed the first phase of a Gulf of Mexico Regional Ecosystem Restoration Strategy to address the long-term degradation of Gulf ecosystems, some of it due to the 60-year history of oil and gas development in the region.

Unfortunately, the US Congress — caught up in partisan rancour, including debates about expanding offshore oil drilling — has failed to adopt legislation to address the lessons learned and the recommendations of the oil-spill commission and others. Such legislation should codify the executive reforms mentioned earlier into law, increase liability limits, and dedicate sustained funding for oil-spill research and environmental assessment and monitoring.

Even in the current constrained fiscal circumstances, improved oversight and essential R&D could be supported by industry fees amounting to pennies per barrel, imperceptible within the daily fluctuations in price on the world market or at the pump.

New laws were passed within a year of the 1989 *Exxon Valdez* spill in Alaska. If important lessons are not to be lost as the events of 2010 fade from memory, there is a pressing need to change the law to make such accidents less likely, and our response more effective. ■

Donald Boesch is president of the University of Maryland Center for Environmental Science in Cambridge.
e-mail: boesch@umces.edu

ROOT CAUSES OF THE
BLOWOUT WERE
**DEFICIENCIES
IN REGULATORY
OVERSIGHT
AND MULTIPLE
POOR DECISIONS.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/6p74jb

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

IMMUNOLOGY

How fat spurs inflammation

Obesity increases the risk of metabolic conditions such as insulin resistance by triggering inflammation. Lan Wu and Luc Van Kaer at Vanderbilt University in Nashville, Tennessee, and their colleagues have established that a subset of immune cells called invariant natural killer T (iNKT) cells links fat build-up with inflammation.

The authors fed mice a high-fat diet, which activated their iNKT cells, triggering the release of inflammatory proteins. When the animals' iNKT cells were stimulated with an injected molecule for eight weeks, the mice developed insulin resistance and fatty livers. By contrast, mice that were engineered to lack iNKT cells did not develop excess inflammation or metabolic conditions on a high-fat diet, despite becoming obese.

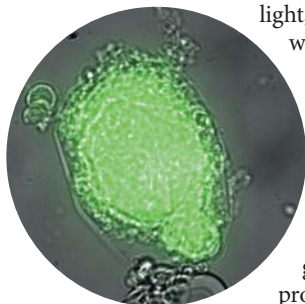
Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1200498109> (2012)

NANOTECHNOLOGY

Injectable protein nanofactories

Tiny injectable particles containing protein 'factories' can be remotely activated by ultraviolet light. The technique could ultimately be used for drug delivery.

Daniel Anderson and his group at the Massachusetts Institute of Technology in Cambridge created lipid spheres more than 100 nanometres in



diameter that contained DNA and all the cellular ingredients and machinery needed to make proteins. By tagging the DNA with a chemical group that prevents it from being transcribed into RNA but can be removed using ultraviolet

light, the researchers were able to control activation of RNA and protein production. The authors developed nanoparticles that produce green fluorescent protein (pictured) and luciferase, an enzyme often used for *in vivo* molecular imaging. They showed that they could switch

on luciferase production remotely after injecting the nanoparticles into mice.

Nano Lett. <http://dx.doi.org/10.1021/nl2036047> (2012)

STEM CELLS

Recipes for making lung cells

The development of methods to transform embryonic stem cells into lung tissue paves the way for models of lung diseases such as cystic fibrosis.

Darrell Kotton at Boston University in Massachusetts and his team used a series of proteins to turn mouse embryonic stem cells into lung progenitor cells. When implanted into a mouse lung

in abundance across the reef, regardless of how susceptible they were to thermal stress and bleaching. These differences in abundance did not follow changes in latitude or temperature.

This flexibility may enable coral reefs to continue functioning as the environment alters with climate change.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2012.02.068> (2012)

Can coral cope with climate change?

Climate change is likely to alter the species composition of coral reefs, rather than wipe out entire reef ecosystems.

Terry Hughes at James Cook University in Townsville, Australia, and his team sampled 132 sites along the full length of the Great Barrier Reef (pictured), spanning 13° of latitude and a range of sea surface temperatures. Of the 12 coral taxa sampled, 11 showed significant differences



AURELIE MOYA

airway that are disturbed in cystic fibrosis.

Cell Stem Cell 10, 398–411; 385–397 (2012)

NEUROPSYCHOLOGY

Poor attention linked to dyslexia

Early problems with visual attention in children seem to be linked to later difficulties in learning to read, a hallmark of dyslexia.

Andrea Facoetti at the University of Padua in Italy and his colleagues tested 96 Italian-speaking children in kindergarten on their ability to pick out specific symbols from an array of others, as a measure of their visual-attention skills. Those who scored below average in visual attention were more likely to show below-average reading skills when they reached first and second grade.

The results suggest that visual-attention deficits, rather than developmental delays in language skills, could predict dyslexia diagnoses in children, and could inspire strategies for early detection and treatment. *Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2012.03.013> (2012)

INFECTIOUS DISEASE

Multiple malaria mechanisms

The considerable variation in malaria symptoms seen between patients can be predicted from the parasite's expression of certain genes and the response of the host's red blood cells to infection.

Red blood cells infected by the malaria parasite *Plasmodium falciparum* often stick to uninfected ones in a process called rosetting. Peter Bull and his co-workers at the Kenya Medical Research Institute–Wellcome Trust Research Programme in Kilifi studied parasites isolated from 131 infected children and found that patients with high levels of rosetting tended to have breathing difficulties.

Children infected with parasites that did not exhibit high rosetting but expressed high levels of a particular subset of 'var' genes were more likely to exhibit impaired consciousness or coma. Var genes encode proteins that sit on the surface of infected red blood cells and specific subsets have previously been linked to severe malaria.

The findings could help to guide treatment decisions or lead to new therapies.

Sci. Transl. Med. 4, 129ra45 (2012)

PHOTONICS

Diamond sparkles with one photon

A diamond diode can produce single photons at room temperature in response to a small electrical current.

Devices that can produce one photon at a time have potential applications in quantum communication and computation, but usually require very low temperatures. Norikazu Mizuochi at Osaka University in Japan and his colleagues say that the key feature of their diode is a diamond layer that contains gaps in the carbon lattice occupied by nitrogen atoms. The researchers theorize that these nitrogen-vacancy complexes naturally capture positively charged 'holes' and electrons from a small electrical current that recombine and emit light.

Nature Photon. <http://dx.doi.org/10.1038/nphoton.2012.75> (2012)

NEUROSCIENCE

Fragile X fixed in mice

A molecule that inhibits a neurotransmitter receptor reverses many of the symptoms in a mouse model of fragile X syndrome, the most common form of inherited mental retardation. The disorder is caused by mutations in the *FMR1* gene that result in overactivity of the mGlu5 receptor.

Lothar Lindemann at

COMMUNITY CHOICE

The most viewed papers in science

BIOELECTRONICS

Snail drives implanted fuel cell

HIGHLY READ
on pubs.acs.org
in March

A fuel cell implanted in a living snail can generate electrical power for several months.

Implanted fuel cells that are driven by glucose generated by their host could one day power medical devices in humans or environmental sensors in animals. Evgeny Katz at Clarkson University in Potsdam, New York, and his colleagues made the electrodes of their fuel cell out of densely packed carbon nanotubes, and attached glucose-oxidizing and oxygen-reducing enzymes to them. The authors then implanted the electrodes into a snail (*Neohelix albolabris*). After decreasing the rate of current extraction to match the snail's slow glucose transport and metabolism, they got continuous electrical output for an hour. The fuel cell remained functional in the snail for several months.

J. Am. Chem. Soc. 134, 5040–5043 (2012)

F. Hoffmann-La Roche in Basel, Switzerland, Mark Bear at the Massachusetts Institute of Technology in Cambridge and their colleagues treated young adult mice engineered to have fragile X syndrome with a new long-acting mGlu5 inhibitor, CTEP, for up to four months. The treatment reversed a range of defects, including hyperactivity, impaired learning and memory, and abnormal neuronal shape. CTEP also corrected the elevated rate of protein synthesis in the brain's hippocampal region, which is characteristic of the disease.

Several drug companies are testing mGlu5 inhibitors in clinical trials.

Neuron 74, 49–56 (2012)

CLIMATE MODELLING

Predicting the Indian monsoon

Climate models can produce more-accurate forecasts of rainfall during India's summer monsoon season than can conventional statistical analyses of sea surface temperatures.

Timothy DelSole and Jagadish Shukla at George Mason University in Fairfax, Virginia, found no significant



statistical relationship between summer monsoon rainfall in India and sea surface temperatures surrounding the country in May of each year from 1960 to 2005. They analysed predictions made by five coupled ocean–atmosphere climate models based on the same temperature data and period, and found that the models were better than the statistical methods at predicting rainfall.

The authors say the models can more accurately forecast the evolution of sea surface temperatures throughout the summer, which influences the monsoon rainfall (pictured). *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2012GL051279> (2012)

► **NATURE.COM**

For the latest research published by Nature visit:
www.nature.com/latestresearch

N. NANU/AFP/GETTY

SEVEN DAYS

The news in brief

POLICY

Controversy in class

School teachers in Tennessee now have licence to critique mainstream scientific theories often attacked by religious and political conservatives, such as evolution and climate change. A bill that had passed both state legislative houses with an overwhelming majority automatically became law on 10 April, after state governor Bill Haslam (Republican) neither signed nor vetoed it. The law's ramifications remain unclear, but opponents worry that it could result in the teaching of religious or non-scientific theories in science classrooms. Tennessee is the second US state to enact a law of this kind; the first was Louisiana in 2008. See go.nature.com/iy5xcd for more.

Antibiotic curb

Spurred by long-running concerns about the rise of resistant strains of bacteria, the US Food and Drug Administration (FDA) announced on 11 April a multi-pronged initiative to cut back on the use of antibiotics in farm animals. Guidance documents recommend that drug firms remove 'growth' and 'production' uses from their product labels; that farmers use certain antibiotics only for therapeutic purposes; and that veterinary surgeons step in as watchdogs. Because these measures are voluntary for the moment, sceptics say that the initiative lacks the punch of robust regulation. See go.nature.com/nwn2nf for more.

Fracking go-ahead

Scientists appointed by the UK government have recommended that hydraulic fracturing, or fracking — in which fluid is pumped



North Korean rocket fails

After much hype, North Korea's latest space shot ended in failure shortly after lift-off on 13 April. Seconds after launch, the first stage of the Unha-3 rocket plunged into the Yellow Sea, 165 kilometres west of Seoul. The United States military, which tracked the launch, says that the rest of the rocket seems to have quickly followed, along with the small weather satellite that it was carrying. The failure is the nation's third since 1998. Unlike with previous attempts, the North Korean government acknowledged the mishap in a brief statement to the press.

into rock to release natural gas — should continue, despite the risk of triggering earthquakes. On 17 April, the panel published a report recommending a preliminary injection stage with near-real-time monitoring of seismic activity. It advises halting fracking after an earthquake of magnitude 0.5 or greater. Last year, a report by Cuadrilla, an energy company based in Lichfield, UK, proposed a threshold of 1.7. Both reports were commissioned following earthquakes in April and May 2011, of magnitude 2.3 and 1.5 respectively, near a UK Cuadrilla fracking well. See go.nature.com/jgciyf for more.

Stem-cell safety

Regulations for the safe clinical use of stem cells and other "investigational agents" approved by the Texas Medical Board on 13 April have been criticized for being too permissive and opening the door to the sale of dubious treatments. They require those who wish to provide stem cells to patients to gain approval either from the US Food and Drug Administration, which requires most stem-cell treatments to be clinically verified, or from an institutional review board. The latter hurdle is likely to be lower. The rules were introduced by Governor Rick

Perry, who had stem-cell treatment for back pain last summer, and approved by a board of Perry appointees. See go.nature.com/owy9or for more.

Emissions up

Greenhouse-gas emissions in the United States rose by 3.2% in 2010, marking the end of a two-year slide that coincided with the global economic crisis, according to an inventory released by the Environmental Protection Agency on 16 April. Cumulative emissions rose to the equivalent of 6.8 billion tonnes of carbon dioxide, driven by fossil fuel consumption in the electricity, industrial and transport sectors. Despite the sharp increase from 2009, emissions remain more than 6% below their 2007 peak. Australia's emissions, meanwhile, rose by 0.6% in 2011, to the equivalent of 546 million tonnes of carbon dioxide, according to government figures released on 17 April.

BUSINESS

Drug fines

Improper marketing of antipsychotic drugs is attracting further fines. Last week, US health-care giant Johnson & Johnson, with its subsidiary Janssen Pharmaceuticals, was ordered to pay the state of Arkansas more than US\$1.2 billion for playing down risks associated with the blockbuster antipsychotic Risperdal (risperidone), including increased probability of death in some trial groups. A string of states are suing the firm, which is headquartered in New Brunswick, New Jersey; fines have so far been ordered in South Carolina (\$327 million) and Louisiana (\$258 million). Johnson & Johnson has appealed both

KYODO/PHOTO SHOT

decisions, but agreed to pay \$158 million to Texas in a settlement. In 2009, Eli Lilly of Indianapolis, Indiana, agreed to pay \$1.4 billion in fines related to its marketing of another antipsychotic drug, Zyprexa (olanzapine).

PEOPLE

Biologist sacked

Protein crystallographer Robert Schwarzenbacher has been fired by the University of Salzburg in Austria following allegations that he fabricated the molecular structure of a version of a birch-pollen allergen, published in the *Journal of Immunology* in 2010 (N. Zaborsky *et al.*, *J. Immunol.* 184, 725–735; 2010). The article has not been retracted but the structure (code 3K78) has been removed from the Protein Data Bank. Schwarzenbacher admitted fabrication at first (N. Zaborsky *et al.*, *Acta Cryst.* F68, 377; 2012), but then retracted his statement. In a further statement to Austrian news agencies, he then said that the structure contains errors but they were not fraudulent. He is fighting his dismissal.

Scientists arrested

The arrest of two scientists in one week by police in Kolkata has triggered outrage in India. Ambikesh Mahapatra, a



chemist at Kolkata's Jadavpur University, was arrested on 13 April after he circulated a political cartoon that made fun of Mamata Banerjee (**pictured**), the first woman to be chief minister of West Bengal. His arrest follows that of Partho Sarothi Ray, a molecular biologist at the Indian Institute of Science Education and Research Kolkata, who was arrested for protesting against the forcible eviction of slum dwellers in East Kolkata. See go.nature.com/blo21v for more.

Astrophysicist death

No criminal charges will be brought over the death of astrophysicist Steven Rawlings of the University of Oxford, UK. Rawlings was found dead at the home of his friend and fellow Oxford academic Devinder Sivia in January (see go.nature.com/kuiphg). Sivia was arrested and released

on bail, but the UK Crown Prosecution Service now says that he will not face any charges. A coroner's inquest will seek to determine the cause of death. Rawlings was a key figure in the Square Kilometre Array radio telescope project.

EVENTS

Sumatra quake

Seismologists were surprised at the power of an undersea earthquake of magnitude 8.6 off the west coast of Aceh on the Indonesian island of Sumatra on 11 April. The quake caused few fatalities, and there was no major tsunami because it was at a 'strike-slip' fault, in which sections of crust rub together horizontally, rather than thrusting over or under each other. It may be the largest known such quake, apart from one of magnitude 8.6 in northeast India and Tibet in 1950. See go.nature.com/nwgc7a for more.

Satellite silence

Officials with the European Space Agency (ESA) say that they have lost contact with Envisat, the agency's premier Earth-observing satellite. Launched in 2002, the satellite is billed as the most sophisticated environmental monitor in orbit, with ten instruments providing streams

COMING UP

22–27 APRIL

At the European Geosciences Union General Assembly in Vienna, researchers discuss the role of geoscientists in warning of and mitigating natural disasters. www.egu2012.eu

26–27 APRIL

US science-policy experts meet in Washington DC to discuss the nation's science and technology prospects at the 37th Annual American Association for the Advancement of Science Forum on Science and Technology Policy. go.nature.com/eub4u7

of valuable data on everything from ozone, clouds and greenhouse gases to land-use trends and sea surface temperatures. Its signal cut out on 8 April, and although the satellite remains in stable orbit around Earth, ESA could not establish contact as *Nature* went to press. See go.nature.com/9h5mom for more.

RESEARCH

China ethics

More than 1,000 Chinese science journals administered by the China Association for Science and Technology in Beijing have jointly declared a code of ethics to prevent academic misconduct. State media reported on 10 April that publishers would punish editors who knowingly published plagiarized articles or abused their positions. The statement comes a month after China's education ministry launched an effort to curb scientific misconduct at universities.

► **NATURE.COM**

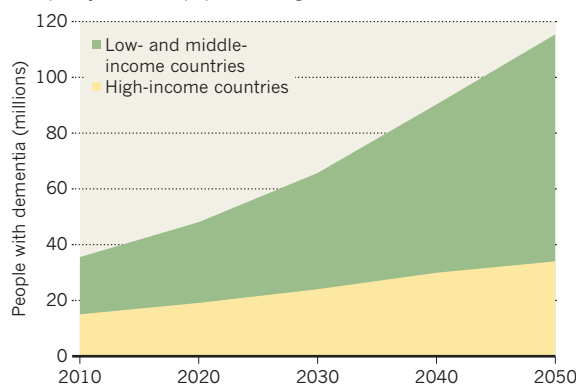
For daily news updates see: www.nature.com/news

TREND WATCH

An estimated 35.6 million people were living with dementia in 2010, according to a World Health Organization report released on 11 April, which urged governments to make plans for coping with the economic and social burden of brain illnesses. As people live longer, cases of dementia are projected to reach 115 million in 2050 (see chart). That estimate assumes current prevalence, which in people over 60 ranges from 8.5% in Latin America to 2.1% in sub-Saharan Africa.

RIISING TIDE OF DEMENTIA

The number of people with dementia is expected to triple by 2050* as populations age.



*Extrapolations based on world dementia prevalence of 4.7% in those aged 60 and over in 2010.

NEWS IN FOCUS

JAPAN Increase to tsunami threat puts towns in a bind **p.296**

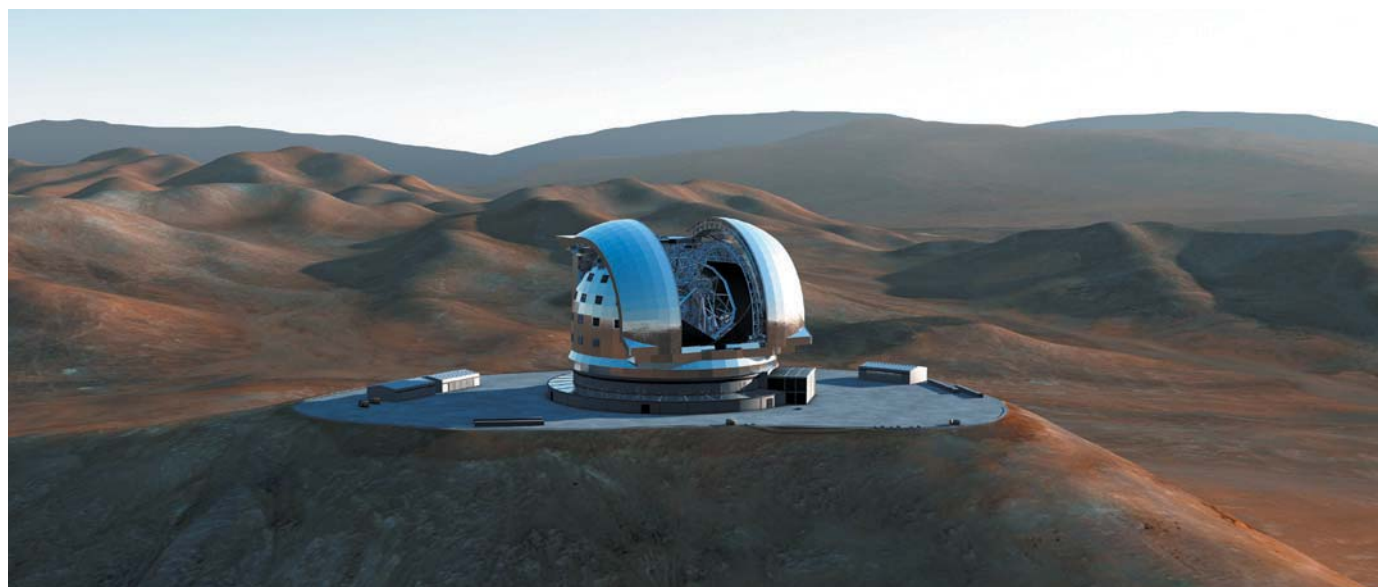
Q&A France's presidential contenders clash over science **p.298**

GENOMICS Sequencing projects hunt for roots of mental disability **p.302**

LAW Underage killers seek a defence in neuroscience **p.304**



ESO



The construction of the European Extremely Large Telescope (artist's impression) in Chile will rely on funds from Brazil.

RESEARCH CONSORTIA

European groups go global

International collaborations seek far-flung members to close budget gap.

BY GEOFF BRUMFIEL

The European Organization for Nuclear Research (CERN) draws particle physicists from around the world to its headquarters near Geneva, Switzerland, but when they line up for lunch the meals are traditionally French. Duck-leg confit and sliced beef liver are among the delicacies on this week's cafeteria menu.

But falafel may soon become a staple. Last month, Israel took its seat on CERN's council as an associate member — the first non-European country to do so in the organization's 50-year history. Other nations outside the European Union, including Serbia and Turkey, are expected to follow soon. For Rolf-Dieter Heuer, the lab's director-general, the goal is to “redefine the ‘E’ in CERN from ‘Europe’ to ‘everywhere’”. CERN is not alone in its internationalist overtures: facilities across the continent are hoisting the flags of new, non-European members.

Part of the motivation for wooing new

partners is a genuine desire to broaden scientific collaborations, but it is no coincidence that the expansion coincides with the worst financial crisis in a generation. Many European nations are slashing their science spending, and international labs are feeling the pressure to find savings and keep costs down¹. Greece, facing a deepening crisis, has sought to reduce its fees to CERN and the European Space Agency (ESA), which currently total about €30 million (US\$39 million). And states such as Portugal and Spain, which have made particularly deep cuts to their science budgets², are almost certain to try to renegotiate their international obligations. Meanwhile, fast-developing countries outside Europe have money to spend on science.

Economics has always played a part in the evolution of Europe's multinational research facilities. Many were set up as international treaty organizations in the 1950s and 1960s

as part of an effort to strengthen economic, social and industrial ties across Europe. They have gradually expanded their membership, but have traditionally drawn a line at the edge of the continent. Although large numbers of US and Japanese scientists work at CERN, for example, those nations are designated as observers on the lab's governing council.

That is changing fast. Last year, India joined the Institute Laue-Langevin (ILL) in Grenoble, France, which produces beams of neutrons to probe the structures of materials and biochemicals. Together with Russia, India has also joined the €1.4-billion Facility for Antiproton and Ion Research (FAIR) now under construction in Darmstadt, Germany. And later this year, Brazil is expected to ratify its membership in the European Southern Observatory (ESO), based in Garching, Germany³.

The financial downturn has certainly been a factor in the decision to bring India in as a full member of the ILL, says Andrew Harrison, the facility's director. The ILL's nuclear reactor ▶

► **NATURE.COM**
For more science stories from Europe, see:
go.nature.com/ptxmzt

► delivers neutron beams to dozens of instruments that have to be regularly upgraded to keep the facility globally competitive. With each upgrade costing between €3 million and €6 million, staying on the cutting edge eats up a significant chunk of the facility's €88-million annual operating budget. India will bring hardware and an extra €1 million or so per year in cash.

CRUCIAL PARTNERS

The desire to fund entirely new facilities provides an even stronger motivation for inviting outside partners. Beatrix Vierkorn-Rudolph, who oversees large research facilities at the German Federal Ministry of Education and Research in Bonn, says that without support from Russia and India, FAIR could not have gone ahead. The facility's Russian-born director, Boris Sharkov, adds that FAIR is in negotiations with China, Brazil and Saudi Arabia about possible collaborations, which could help to fund new particle detectors.

Brazil's admission into ESO will be crucial in paying for that organization's next-generation telescope, the 40-metre European Extremely Large Telescope (E-ELT). Around one-third of the €1.1-billion construction budget will come from Brazil's membership dues, says Tim de Zeeuw, director-general of ESO.

The new members get plenty in return, benefiting from the prestige associated with international collaborations and the training offered to their scientists. "ESO facilities are the top of the line," says Sergio Machado Rezende, a physicist at the Federal University of Pernambuco in Recife, Brazil, and the country's science minister from 2005 to 2010. ESO's best telescopes are located in nearby Chile, which made joining the organization even more attractive, he says. Brazil's astronomy community is currently made up of only about 200 researchers,

but Rezende says that involvement in ESO will help the community to grow and gain strength. Brazil's robust economy means that the nation will probably also consider involvement in organizations such as CERN, he says.

India's membership of FAIR should provide benefits for the country's technology firms, says Subhasis Chattopadhyay, a scientific officer at the Variable Energy Cyclotron Centre in Kolkata. The companies will supply power converters and magnets for the facility, and so gain expertise in constructing these hi-tech instruments. "Our goal was to get Indian industry building many new things," says Chattopadhyay, adding that India, too, has its eye on CERN.

Not every organization is looking to expand beyond Europe. "For the time being, there are really zero discussions," says Frederic Nordlund, acting head of international relations at ESA in Paris. His agency works closely with many nations, including Russia and India, but most want to maintain domestic expertise in space

"If you get the best players, then you develop the best facilities."

only new European Union member states.

The European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, stands apart for different reasons. "EMBL isn't an organization that grows by building mega-projects," says Iain Mattaj, director-general of the organization. As a result, it has no plans to open itself formally to outside membership — although it did allow Australia to join as an associate member in 2008.

Those that are expanding recognize that the move comes with risks. Some fear that a growth in international membership could erode the

quality of European science or make it more difficult to govern. But Harrison says that European nations already on the ILL council are careful to ensure that any new members have a strong scientific community ready to use the facility. "We would not be interested in a member country that is rich but did not have the kind of user base that we could work with," he says.

SUSHI SURPRISE

Things could get complicated if Europe's centralizing instincts were to lead it to consolidate facilities — merging ESA and ESO, for example, or combining all of Europe's particle-physics facilities. Although such mergers are not likely in the short term, the presence of many non-European partners would hamper them in the future, says Carlo Rizzuto, chairman of both the board of the Elettra Laboratory near Trieste, Italy, and the European Association of National Research Facilities. But ultimately, Rizzuto thinks, more international members will only strengthen Europe's facilities, and its science. "If you get the best players, then you develop the best facilities," he says.

Heuer hopes that the CERN council's membership will grow from 20 to around 26 in the next 5–6 years. At the same time, he stresses that the lab must maintain its European character of careful deliberation and consensus-building, which is one reason that new partners must first join as associate members. Heuer predicts that CERN will adapt smoothly, given that it already hosts many scientists from outside Europe. As he points out, the lab's cuisine has already begun to diversify: "The cafeteria also sells sushi, by the way." ■

1. Brumfiel, G. *Nature* **466**, 1028–1029 (2010).
2. Catanzaro, M. *Nature* <http://dx.doi.org/10.1038/nature.2012.10242> (2012).
3. Mann, A. *Nature* **469**, 451–452 (2011).

DISASTER PREPAREDNESS

Tsunami simulations scare Japan

Updated risk assessment warns country to prepare for much larger earthquakes and waves.

BY DAVID CYRANOSKI IN TOKYO

Japan's government is heeding a key message from last year's Tohoku earthquake and tsunami: the underwater faults that encircle the country can unleash much greater devastation than previously anticipated. Last week, the cabinet's disaster-management division briefed local officials on simulations that raise the spectre of waves even larger and more destructive than those last March, sending the officials scrambling to rethink their tsunami defence plans.

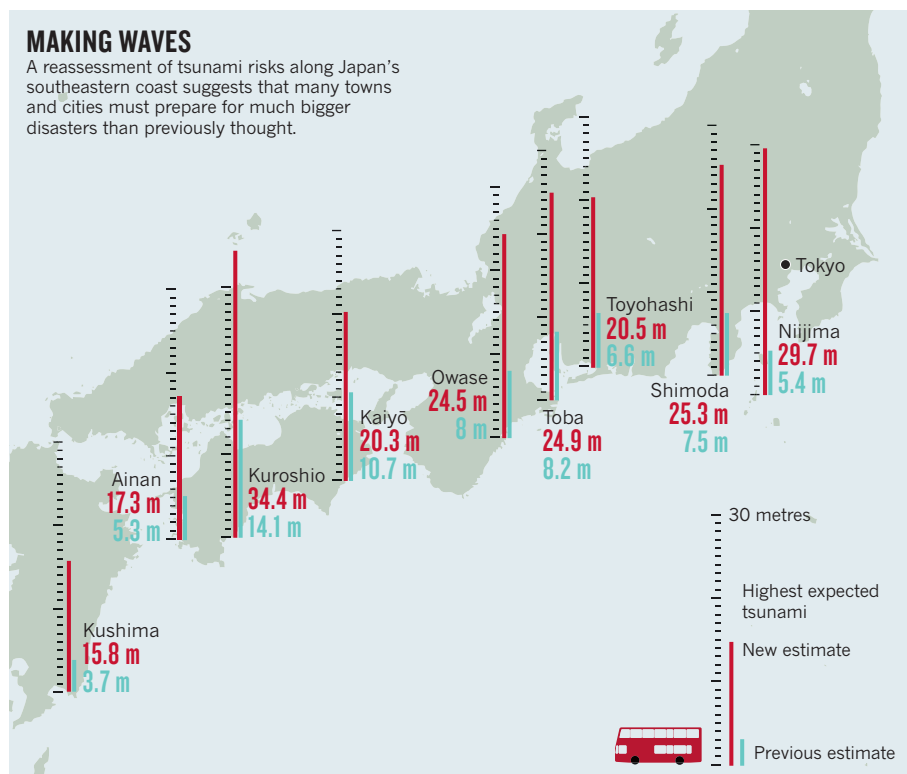
The estimates come from a government-appointed team of scientists led by Katsuyuki Abe, a tsunami expert and emeritus professor of the University of Tokyo. The team updated the nation's main tsunami model by increasing the magnitude of the largest expected earthquake to incorporate recent massive quakes, including the magnitude-9.0 Tohoku and the magnitude 8.8 that hit Chile in 2010, causing a tsunami that reached as far as Japan. The result: for most locations, the size of future tsunamis could far outstrip previous estimates, which were made

in 2003. The town of Kuroshio in Kōchi prefecture was steered for a maximum 14.1-metre wave; it now faces the threat of a 34.4-metre inundation. Likewise, the offshore island of Niiijima has to prepare for a 29.7-metre wave, rather than 5.4 metres (see 'Making waves').

Ninety cities and towns must consider how to withstand a tsunami wave of higher than 10 metres, whereas just ten locales were expecting this from the older simulations. Twenty-three have been told to prepare for a tsunami of 20 metres or more, a threat none had previously

MAKING WAVES

A reassessment of tsunami risks along Japan's southeastern coast suggests that many towns and cities must prepare for much bigger disasters than previously thought.



anticipated. The team also raised the estimates of the risks posed by the largest earthquakes, with the number of towns and cities expecting the maximum level of ground shaking — 7 on Japan's intensity scale — raised from 35 to 153.

The reappraisal focuses on the Nankai trough, an offshore fault south of Honshu, Japan's main island, that regularly produces large quakes, including the magnitude-8.7 Hōei quake of 1707 that set the ceiling for the 2003 estimates. The latest simulations use the Tohoku quake, which was roughly three times more powerful than Hōei, as the maximum. They also take account of recent studies that used the thickness of sediment layers deposited by past tsunamis to estimate the size and frequency of major earthquakes and tsunamis that happened before accurate measurement methods existed. The 2003 models, which predicted much smaller tsunamis than those that followed the Tohoku earthquake, have been criticized for not including such sedimentation data (see *Nature* **483**, 141–143; 2012).

The updated model also considers 'large slip' areas, in which extensive crust movement can make parts of tsunamis particularly hefty. Such areas were to blame for the unexpectedly high waves in parts of Japan last year. The team considered how coastal regions would be affected in 11 tsunami scenarios, each of which had large slip areas in different places along the Nankai trough.

Magnitude-9 events are expected to occur very rarely, perhaps once every millennium, says Kenji Satake, a tsunami expert at the University of Tokyo and a member of the team that updated the models. How much preparation

is needed for such rare, devastating events remains a matter of debate, but the models have put local governments in a tight spot. Yukihiro Nakamura, head of earthquake preparedness in the Kōchi prefecture, says that the briefing was not detailed enough to give administrators a clear path forward.

"We need to know how high the water will come at different points. Will it be 34 metres everywhere?" The government has promised a fine-grained simulation and inundation maps that detail water levels by the end of May.

Even that information may not help officials to determine how to prepare for such devastating events. Some towns are installing additional loudspeakers to warn citizens of an approaching tsunami, but they might give no more than a few minutes' warning. Nakamura says it is unreasonable to think that people could seek shelter — by climbing evacuation towers tens of metres tall, for example — within that time.

Officials in some regions are also making plans to move city offices to higher ground, and considering measures such as building large underground shelters or relocating large numbers of homes — both of which are costly options that are beyond the capacity of local or prefectural governments. A cabinet working group is discussing how the central government could help.

Before last year's devastating earthquake, "we had been preparing for a magnitude-8 quake they said would have 60% chance of coming in the next 30 years," says Nakamura. "But after Tohoku, we have to be prepared for the unexpected, too." ■



Q&A The French election

A question of science

French scientists will have the chance to give their verdict on the root-and-branch reforms to the country's science system on 22 April, when they vote in the first round of France's presidential elections. Polls predict a defeat for Nicolas Sarkozy, the incumbent president and candidate for the right-wing UMP party. That result would cheer those researchers who argue that Sarkozy's sweeping reforms, often imposed with little consultation, have left some institutions sidelined and many labs lacking funds. Others, however, support many of his research reforms, including his move to give universities more autonomy, and fear that the programme could stall if he loses power.

Sarkozy's main rival, the Socialist Party's François Hollande, has consistently led in the polls and is predicted to win the second and final round on 6 May by a large margin. If he does, he will be the first socialist to take the Élysée Palace since François Mitterrand was re-elected in 1988. Hollande says that he will maintain the thrust of many of the major reforms begun by Sarkozy, but will ensure greater consultation with the research community as he takes them forward. He also promises to create many more science jobs, in particular for young researchers. Although none of the other eight candidates in the first round is likely to win, horse-trading for the transfer of votes in the second round could influence the policies of the eventual victor.

Nature asked the two leading candidates about their plans for science; their full responses, in French and English, are available at go.nature.com/mjvmu4, with extracts appearing here.

NICOLAS SARKOZY

What have been the main accomplishments of your presidency?

Without hesitation, the 2007 law giving universities autonomy from centralized state control. It has given universities greater freedom to set their own strategies, allows them to own their land and buildings, manage their budgets and borrow capital to make major investments. The universities can now recruit scientists without any ministry involvement — even though researchers are civil servants — and they have greater freedom to set salaries, through bonuses, instead of being tied to rigid national salary scales. Since 2007, university funding has also increased by 22%, and in 2008 we launched a €5-billion [US\$6.5-billion] plan to refurbish a dozen select university campuses.

Some scientists say that excellence initiatives are creating a two-tier system. How will you maintain research and teaching across France?

Idex, the initiative for excellence funded by €7.7 billion from the €35-billion Investments for the Future programme, is exactly analogous to an initiative launched by Germany several years ago that has seen the emergence of nine institutions destined to become world-class universities. We have similarly selected eight.

FRANÇOIS HOLLANDE

French universities have experienced huge changes under President Sarkozy. Would you continue or change his programme?

The 2007 law on university autonomy must be profoundly reformed, but there will be no going back on the principle of autonomy. However, when responsibilities are transferred from the state to the universities, there needs to be a comparable transfer of resources.

I want the university-funding mechanisms to be clarified, and for these not to increase disparities, in particular by cutting back on teaching or the salaries of lecturers and full-time researchers. The reforms will be discussed within a framework law on higher education and research at the end of 2013, preceded this year by a national consultation.

Many leading scientists complain that lab funding by the national research organizations is falling. How would you rectify this situation?

This year and last, the overall funding levels for the national research organizations and the National Research Agency fell. This proves that the President's words are just smoke and mirrors. I will rebalance core lab funding — the permanent support that lets teams think in the medium and long term —

There is not that much disparity in opportunities among the universities. They have other opportunities beyond Idex. Ninety per cent of French universities have received some form of support from the Investments for the Future programme. If we want world-class research champions, we need to concentrate resources on those places where the research is best.

I'm a big fan of cycling, and I love watching the Tour de France. We've never seen the pack accelerate because those at the rear go faster; the pack accelerates when the leaders accelerate. Similarly, I think the Investments for the Future programme, because it rewards the best, will lift the entire French system of research and higher education, not just the winners.

What is the future for the large national research agencies, such as the CNRS and INSERM?

France is a very special case. Just after the Second World War, we created agencies separate from the universities to do basic research. At the time, such a set-up was found only in Communist countries, in particular the USSR and China. Now, even these countries have abandoned this model. In the United States, the vast majority of research operators are universities. That is not yet the case in France, but it is our goal. The research organizations are expected to become more like real funding agencies, serving the universities and research institutions.

Many leading scientists complain that lab funding by the national research organizations is falling. How would you rectify this situation?

In 2011, research spending as a percentage of gross domestic product reached 2.3%, close to its historical high [of 2.36%] in the early 1990s. The budgets of the main public

"We need to concentrate resources on those places where the research is best."

research agencies — the CNRS, INSERM, CEA, INRA — have risen every year since 2007. The National Research Agency's budget has stagnated, but the agency is also responsible for allocating €22 billion for higher education and research under the Investments for the Future programme. The best French research teams have enjoyed an unprecedented increase in funding, but these funds have not necessarily passed through the main research agencies.

Scientists say that you have increased bureaucracy in an already-complex research landscape. How would you streamline this?

It is true that new structures have emerged. Many need to be consolidated, and others need to be simplified or reduced. The landscape of French research and higher education is being restructured. It is like a major construction site

where, as the project advances, the scaffolding is taken down, the tarpaulin is taken off and the new building appears. That is how it will be for the universities and research in France.

Would you maintain nuclear power's 75% share of French electricity generation?

I want to intensify research to develop safer nuclear generators, but also boost research and development into alternative energy sources. I want to manage the inevitable transition towards alternative energies that will take place in France. In the meantime, it would be stupid and counterproductive to deprive France of its valuable nuclear-energy assets.

Would you change France's laws on human embryo and embryonic-stem-cell research?

The issue was widely debated in parliament during the revision of the bioethics laws in 2011. The current situation, in which embryo research is prohibited but there are dispensations for work that could lead to major therapeutic progress and where alternatives are not available, is strictly equivalent from a scientific point of view to what the Socialists propose. During the 2011 debate on bioethics, the French overwhelmingly expressed their commitment to the sanctity of life from its beginning. The national representation therefore ruled that the existing system should be maintained.

with project funding. I will refocus the National Research Agency [the main source of project funding] on national priorities, emerging themes and interdisciplinary projects. I will order a financial audit of the plans launched by the previous government.

Other countries think French research is being inundated with money. But this new money doesn't compensate for the cuts observed at the CNRS over the past two years, which has resulted in 25% less cash in the labs. There is a chasm between the €22 billion announced for Investments for the Future and the reality, which is that the funding is only the interest on that endowment distributed over ten years — for every €1 billion announced, the reality is just €34 million per year.

Some scientists say that excellence initiatives are creating a two-tier system. How will you maintain research quality across France?

We need to restore the true sense of the term 'excellence'. The Investments for the Future programme has worsened disparities between universities. French research is not just Paris, Bordeaux, Toulouse, Aix, Marseilles and Strasbourg. I will not reconsider decisions that have been made but, after the audit, I will demand that unallocated credits be used to correct geographical disparities, to ensure that these do not develop into scientific and university

deserts. I will push to replace competition between universities with cooperation.

Would you prioritize research spending?

I will make higher education and research a priority, because the major problem facing France is to prepare for the future. I will make youth the core of my programme, giving priority to education from early childhood through to university. I will favour basic research, which has been sacrificed these past few years, paying particular attention to the life sciences, where funding is far below the international average, and environmental sciences.

"The trust between researchers and political leaders has been broken."

What is the future for the large national research agencies, such as the CNRS and INSERM?

We shouldn't break what works. France is lucky to have its large national research organizations alongside the universities. These organizations must recover their role of steering the national research effort, and of supporting research carried out in the universities.

What will be your policies on nuclear and other energy technologies?

I want France to accomplish an energy

transition with the share of nuclear energy being reduced from 75% to 50% of electricity production within 15 years. I will create major research programmes on 'disruptive' technologies such as electricity storage, the greening of various production sectors and energy efficiency.

Would you change France's laws on human embryo and embryonic-stem-cell research?

I will authorize research on embryonic stem cells to clarify the situation of the 2011 law on bioethics, which maintained a ban on this research while allowing dispensations for particular projects. It is important to put an end to this hypocrisy.

How do you assess President Sarkozy's impact on research and universities?

The President has introduced multiple reforms without providing the necessary resources. In the past decade, we have fallen from 4th to 15th place worldwide in terms of spending on research as a percentage of GDP — 2.26% compared with 2.82% in Germany and 2.9% in the United States. France is trailing, and the trust between researchers and political leaders has been broken. My first objective will be to restore that trust. ■

INTERVIEWS BY DECLAN BUTLER

TECHNOLOGY

Japan gambles on displays

Consortium of three big brands gains government backing to spur mobile technologies.

BY KATHERINE BOURZAC

With ever more beautiful and addictive smartphones and tablets captivating consumers, the market for mobile display screens is booming. Yet Japanese display makers Hitachi, Sony and Toshiba, once industry giants, have languished. Now the three companies have merged their mobile-display subsidiaries to form Japan Display Incorporated, launched on 1 April. With ¥150 billion (US\$1.85 billion) in research and development funding over the next five years, the Tokyo-based endeavour aims to launch technologies that will help its members to recapture the display market.

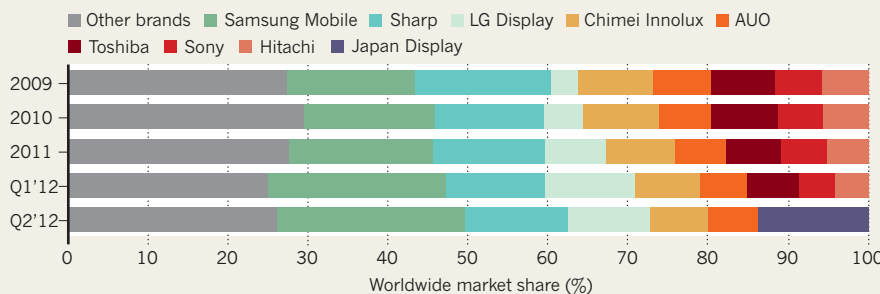
"We expect this company will be driven by research and development," says Hiroyuki Oshima, Japan Display's deputy chief technology officer. Many of the features that sell tablets and smartphones are based on display innovations, such as the super-high-resolution screens of Apple's iconic iPad and iPhone devices. Now the race is on to lengthen battery life, make gadgets thinner and lighter, and provide rich colours that can be seen from wide angles for watching videos with a friend. "The display technologies that have the slickest stuff command the biggest market," says Nick Colaneri, head of the Flexible Display Center at Arizona State University in Tempe.

Japan Display's bid is mainly financed by a government-backed investment fund, the Innovation Network Corporation of Japan, which has contributed ¥200 billion and holds a 70% share in the company. A particular area of focus, Oshima says, will be organic light-emitting diodes (OLEDs), which the company will begin producing by the end of next year. OLEDs, which use polymers or smaller carbon- and metal-based molecules, provide

"Aggressive investment in research and development could make a difference."

A DISPLAY OF UNITY

Toshiba, Sony and Hitachi hope to reverse a steady decline in their share of the mobile display-screen market by combining forces to form 'Japan Display', with a market share that will be the second largest worldwide.



SOURCE: DISPLAYSEARCH

richer colour than conventional liquid-crystal displays (LCDs), and use less power.

Stephen Forrest, vice-president of research at the University of Michigan in Ann Arbor, notes that OLEDs are a speciality of South Korea's Samsung Mobile Display, the world leader with 23.3% of the mobile-display market. Yet, he adds, there is still room for innovation. "Some things about OLED technology are not settled," he says. For example, researchers are still figuring out the most cost-effective design for the electronics that drives the displays. Low-cost 'amorphous-silicon' transistors are adequate for conventional LCD screens, whose pixels change their light-transmitting properties depending on an applied voltage. By contrast, OLED pixels are driven by changes in current, and amorphous-silicon transistors cannot provide enough. One solution is to subject amorphous silicon to a laser treatment that upgrades it to better-quality polysilicon. But this is an expensive process.

The leading alternative to polysilicon for OLEDs is a new class of transistor material, based on metal oxides. These materials promise good performance at lower manufacturing costs than polysilicon. Before the merger, the three companies were all developing metal oxides, and Oshima says that he is

now trying to get the research teams to work together. However, Japan Display may have some catching up to do: on 13 April, Sharp, another Japanese maker, announced that it will become the first to bring metal oxides into production.

In other areas, Oshima says, the research strengths of the three merged companies are more complementary. For example, Sony has developed a technology for embedding touch-screen sensors, typically a separate element sold by another company, into the display itself. If it can be made to work, this 'embedded' touch should make phones thinner and lighter, and reduce costs.

Analysts at DisplaySearch, a market research company headquartered in Santa Clara, California, predict that Japan Display will become a big player, starting with a projected 13.8% market share in its first quarter (see 'A display of unity'). Prior to their joint effort, the three companies didn't invest enough resources in innovation in the mobile-display market, says Oshima. Now, Japan Display will put all its research efforts there. Colaneri thinks that the more focused approach could help the company to grow further. "Aggressive investment in research and development could make a difference," he says. ■



MORE ONLINE

TOP STORY



Karakoram glaciers gaining ice despite warming climate go.nature.com/lyz3hf

MORE NEWS

- Home-grown labware made with three-dimensional printer go.nature.com/dt3yx6
- Infant galaxy offers tantalizing peek at early Universe go.nature.com/azbtmo
- Cash reward for saving Australia's water go.nature.com/lmg7fy

PODCAST



How to minimize the risk from space weather and tune into one voice at a party go.nature.com/scn4lp



Exome sequencing could help to identify the causes of intellectual disability in children such as Siebe.

MEDICAL GENETICS

Gene hunt is on for mental disability

Pioneering clinical genome-sequencing projects focus on patients with developmental delay.

BY EWEN CALLAWAY

Medical geneticists are giving genome sequencing its first big test in the clinic by applying it to some of their most baffling cases. By the end of this year, hundreds of children with unexplained forms of intellectual disability and developmental delay will have had their genomes decoded as part of the first large-scale, national clinical sequencing projects.

These programmes, which were discussed last month at a rare-diseases conference hosted by the Wellcome Trust Sanger Institute near Cambridge, UK, aim to provide a genetic diagnosis that could end years of uncertainty about a child's disability. In the longer term, they could provide crucial data that will underpin efforts to develop therapies. The projects are also highlighting the logistical and ethical challenges of bringing genome sequencing to the consulting room. "The overarching theme

is that genome-based diagnosis is now hitting mainstream medicine," says Han Brunner, a medical geneticist at the Radboud University Nijmegen Medical Centre in the Netherlands, who leads one of the projects.

About 2% of children experience some form of intellectual disability. Many have disorders such as Down's syndrome and fragile X syndrome, which are linked to known genetic abnormalities and so are easily diagnosed. Others have experienced environmental risk factors, such as fetal alcohol exposure, that rule out a simple genetic explanation. However, a large proportion of intellectual disability cases are thought to be the work of single, as-yet-unidentified mutations.

Scientists estimate that about 1,000 genes are involved in the function of the healthy brain. "There are so many genes that can go wrong and give you intellectual disability," says André Reis, a medical geneticist at Erlangen University Hospital in Germany. Reis's group,

the German Mental Retardation Network, has already sequenced the exomes — the 1–2% of the genome that contains instructions for building proteins — of about 50 patients with severe intellectual disability.

Joining the hunt is a UK-based programme called Deciphering Developmental Disorders, which expects to sequence 1,000 exomes by the year's end, with an ultimate goal of diagnosing up to 12,000 British children with developmental delay. A Canadian project called FORGE (finding of rare disease genes) aims to sequence children and families with 200 different disorders this year. And in the United States, the National Human Genome Research Institute in Bethesda, Maryland, recently funded three Mendelian Disorders Sequencing Centers that will apply genome sequencing to diagnosing thousands of patients with a wider range of rare diseases, including intellectual disability and developmental delay.

FIRST GLANCE

Early results are coming in from Brunner's team, which has already sequenced about 100 exomes of children with intellectual disability. By comparing the children's exomes with those of the parents, the researchers have identified new mutations — potential causes of the disorder — in as many as 40% of the cases. The other programmes are having similar success at making possible genetic diagnoses.

In most cases, identifying mutations will not point to medical treatments, let alone cures. But scientists say the importance of a diagnosis should not be discounted. "Parents have been struggling with the delay of their children for years. They have gone from one doctor to the next, had all kinds of tests done on their children looking for an explanation," Reis says. Knowing that the mutation causing a child's intellectual disability is new rather than inherited can also reassure parents that other children they conceive are unlikely to have the same disease.

Treatments could eventually follow. The projects are guiding research in mice, zebrafish and fruitflies, with the goal of unpicking the mechanisms of mental disorders. But it will undoubtedly be a long time before any potential therapies are tried in humans: an early-stage clinical trial of a drug to treat fragile X syndrome, for example, was published last year (S. Jacquemont *et al. Sci. Transl. Med.* 3, 64ra1; 2011), some two decades after the gene underlying the condition, *FMRI*, was identified.

The work is also throwing up a fresh challenge: how can scientists be sure that a specific mutation is the cause of a particular form of mental disability? "It's not clear what is the threshold of evidence at which you can say this is the causal variant in this patient," says

➔ NATURE.COM
Read more about
'genomes on
prescription' at:
go.nature.com/yv7rls

Daniel MacArthur, a geneticist at Massachusetts General Hospital in Boston. In a recent *Science* paper, his team estimated that the average healthy genome contains about 100 gene-disabling mutations. Such 'background noise' could lead scientists astray in their hunt for causal mutations.

Brunner says that about half of the mutations his team has identified have previously been seen in other patients with similar forms of intellectual disability, offering enough assurance to make a diagnosis. Circumstantial evidence, such as indications that the mutation disrupts a gene expressed in the brains of animals, ties the other half of the mutations to intellectual disability. But making a solid case often requires identifying second, third and fourth patients with similar mutations and symptoms.

Scientists are already forging these connections on an informal basis. At the Sanger Institute meeting, several groups reported mutations in a gene called *ARID1B* in patients with intellectual disability. James Lupski, a medical geneticist at Baylor College of Medicine in Houston, Texas, says that when his team identifies a potentially disease-causing mutation in a patient genome, he e-mails other scientists to see whether they have found similar mutations.

But researchers agree that they need a more

formalized way to make these connections. To that end, the US National Center for Biotechnology Information in Bethesda is developing a database, ClinVar, to integrate clinical and genetic data; others, such as DECIPHER, run by the Sanger Institute, handle genetic data such as chromosome rearrangements that can disrupt genes.

The first clinical sequencing projects are also grappling with what they should or shouldn't tell patients. "We don't want people coming into our clinic for intellectual disability and

"Genome-based diagnosis is now hitting mainstream medicine."

coming out with a cancer gene; this is not what they came for," says Reis.

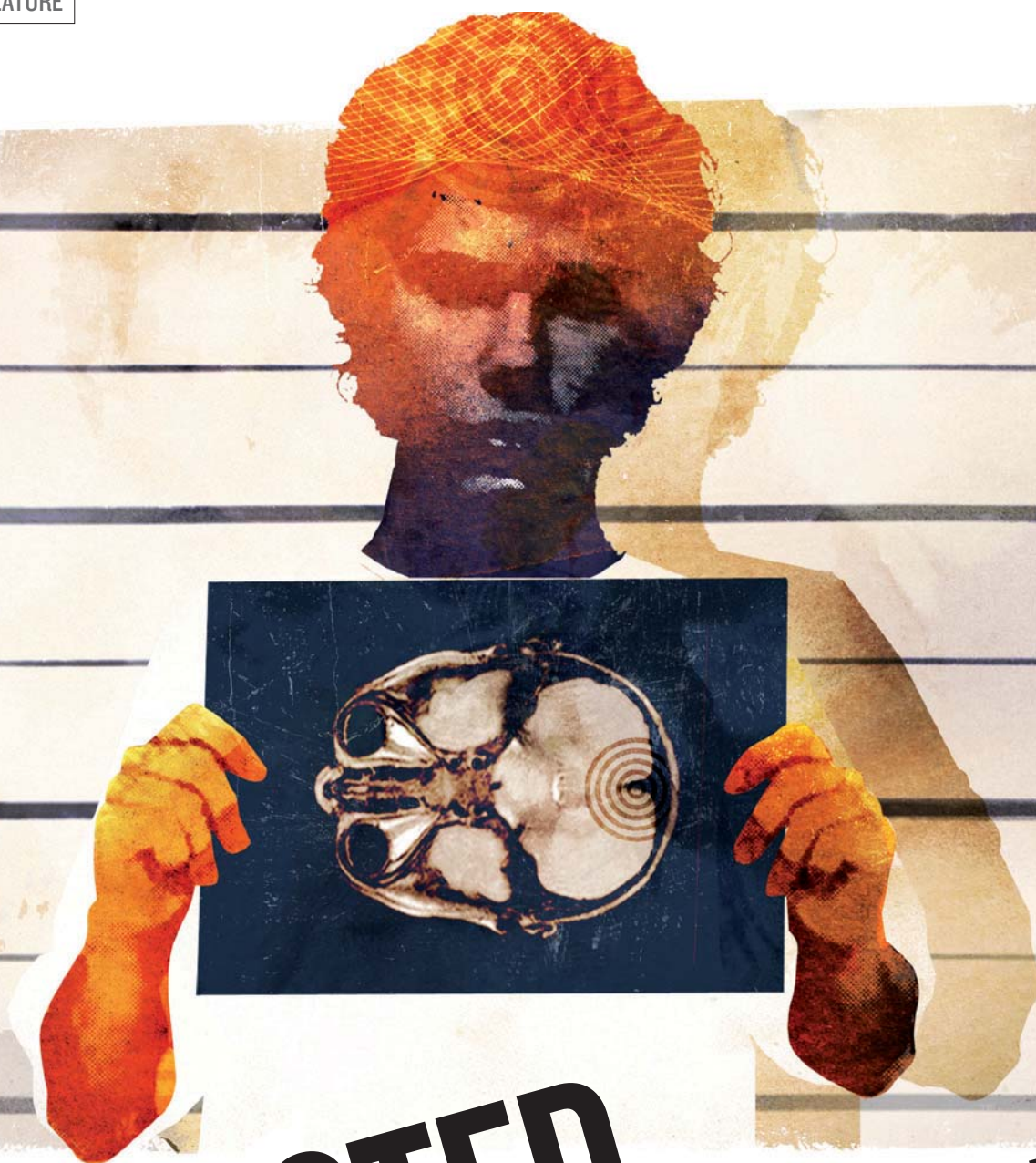
Brunner's team once had to face just that situation. The researchers identified a mutation in a gene in one patient that could increase the risk of colon cancer as an adult. The project's ethical review board had determined that if families wanted to know of mutations potentially underlying a child's intellectual disability, they must also be willing to receive such incidental findings, and so the child's parents were told. But clinical sequencing projects vary in their approach to incidental results. For the time being, Deciphering Developmental Disorders will not inform

families about such findings. For FORGE Canada, the policy varies from province to province.

A working group convened by the American College of Medical Genetics and Genomics in Bethesda recently suggested drawing up a list of gene mutations that ought to be routinely reported back to patients. The list would include mutations strongly linked to conditions for which a medical intervention is available.

"This is a fast-changing ethical environment," says Matt Hurles, a geneticist at the Sanger Institute and one of the leaders of Deciphering Developmental Disorders. His team is conducting a web-based survey to gauge the attitudes of parents, physicians and the general public towards disclosing incidental genomic findings. Lupski admits, "We're learning as we go along. People don't want to hear that, but that's the truth of the matter."

Scientists and clinicians hope that the lessons learned in these initial large-scale clinical sequencing projects will inform genomic medicine as it reaches more patients and moves to other specialties, such as neurology and cardiology, and even to routine health care. "If in five years time this project hasn't catalysed the adoption of genomic technologies which have been shown to be useful, in some degree we will have failed," says Hurles. ■



ARRESTED DEVELOPMENT

Neuroscience shows that the adolescent brain is still developing. The question is whether that should influence the sentencing of juveniles.

BY LIZZIE BUCHEN

ILLUSTRATION BY ANDY POTTS

Teenagers can do terrible things. In 1999, Kuntrell Jackson, then 14, was walking with his cousin and a friend in Blytheville, Arkansas, when they decided to rob a local video store. On the way there, his friend, Derrick Shields, revealed that he was carrying a sawn-off shotgun in his coat sleeve. During the robbery, Shields shot a shop worker in the face, killing her.

Four years later, 14-year-old Evan Miller and an older friend were getting drunk and stoned with a middle-aged neighbour in a trailer park in Moulton, Alabama. A fight broke out, and Miller and the friend beat the neighbour with a baseball bat. Then they set fire to his home and ran, leaving him to die.

Both Miller and Jackson were found guilty of homicide and sentenced to life without parole, meaning that both will spend the rest of their lives in prison. They are not alone. The United States currently has more than 2,500 individuals serving such sentences for crimes they committed as juveniles — that is, before their eighteenth birthdays. It is the only country that officially punishes juveniles in this way. Both Miller and Jackson appealed, arguing that their immaturity at the time of the crime rendered them less culpable for their actions than adults, and that they deserved a less severe punishment. The Supreme Court heard arguments in *Miller v. Alabama* and *Jackson v. Hobbs* in March, and is expected to deliver its ruling by this summer.

The cases are notable not only because they could abolish life-without-parole sentences for juveniles, but also because neuroscience research may play a part in the decision. Several organizations submitted briefs to the court detailing the growing body of research showing that the brain's development continues into at least the early twenties. This, they argue, may explain why young people are more impulsive than adults, more readily swayed by their peers and less likely to consider the consequences of their actions.

UNDER THE INFLUENCE

Advocates for juveniles have been embracing this work as part of a long-term strategy to ensure that young criminals are given less punishment than adults and more opportunities for rehabilitation. And many neuroscientists studying the adolescent brain are gratified that their work is contributing to these efforts. "It's so satisfying to think that maybe in some minuscule way my work was relevant to society," says Bea Luna, who studies adolescent brain development at the University of Pittsburgh in Pennsylvania.

But the brain research may not have as great an influence in court as some scientists and advocates like to think. Some say that the neuroscience offers no fresh insight into adolescent behaviour, and may serve merely as a rhetorical flourish in judges' opinions or as a tool that lawyers and advocates exploit to make their case.

"The neuroscience is being used for an advocacy position," says Emily Murphy of Stanford University in California, who was a fellow with the MacArthur Foundation's Law and Neuroscience Project. "That's all it's always been, in a legal context." Murphy and others worry that the neuroscience currently being used in court may be abused, and might overshadow other research that could make a deeper impact on juvenile crime and punishment.

Historically, courts in the United States have treated juveniles with leniency. But in the late 1980s and early 1990s, a rise in violent youth crime, including several high-profile school shootings, provoked a backlash of tough

A TEENAGER IS LIKE A CAR WITH A GREAT ACCELERATOR BUT TERRIBLE BRAKES.

justice. Across the country, young offenders were increasingly transferred from the juvenile courts, which emphasize rehabilitation, to adult criminal courts, which focus on retribution and punishment, including the death penalty.

Advocates for juveniles have fought to reverse this trend, and have done so, in part, by pointing at major changes now known to take place in the adolescent brain. During maturation, the brain undergoes large-scale structural changes. Fatty tissue called myelin wraps around neurons, speeding up signal transmission in brain cells, particularly between brain regions. At the same time, superfluous connections are pruned away in a process that may decrease noise in the system and allow remaining neurons to function more efficiently.

These processes were once assumed to be completed during childhood, but techniques developed over the past two decades have shattered that idea. Structural magnetic resonance imaging (MRI) shows that pruning goes on into early adulthood. Diffusion tensor imaging, another form of MRI, shows much the same for myelination. And functional MRI (fMRI), which reveals activity in the brain in near real time, has shown how this ongoing development affects activity levels in different parts of the brain. Some of the most marked changes during these later developmental periods occur in brain systems associated with impulse control^{1,2}, resisting immediate rewards³ and emotional processing^{4,5} — all behaviours relevant to criminal activity.

Systems involved in processing reward also seem to mature more quickly than those involved in decision-making and impulse control. One fMRI study⁶ showed that when adolescents were presented with rewards for

performing a simple task, the nucleus accumbens — part of the brain's reward system — lit up in a pattern similar to that seen in adults in the study, but much more strongly. Meanwhile, the sparse activity seen in the prefrontal cortex — a region thought to be involved in decision-making and impulse control — looked more akin to that of children. The authors of the study say that this may help to explain the heightened risk-taking behaviour common in adolescents. A teenager, as advocates and some scientists like to say, is like a car with a great accelerator but terrible brakes. As Larry Steinberg, a psychologist who studies teen behaviour at Temple University in Philadelphia, Pennsylvania, has said, "With powerful impulses under poor control, the likely result is a crash."

CRIME AND PUNISHMENT

The Supreme Court got its first taste of this research in 2005, when it considered *Roper v. Simmons*. In 1994, Christopher Simmons was sentenced to death in Missouri after he and an accomplice wrapped a woman's head in duct tape, bound her limbs with electrical wire, and threw her off a railway bridge. He was 17 at the time of the crime. When the case reached the Supreme Court, the American Medical Association (AMA) and American Psychological Association (APA) filed briefs in Simmons's support explaining the current state of research on the immaturity of adolescent brains.

The court found it unconstitutional to impose the death penalty for crimes committed by juveniles; Simmons is now serving life without parole. Although the neuroscience evidence was never directly cited, Justice Anthony Kennedy wrote in the majority opinion that "as any parent knows and as the scientific and sociological studies ... tend to confirm, '[a] lack of maturity and an underdeveloped sense of responsibility are found in youth more often than in adults'".

Many scientists and advocates for juveniles considered the result a triumph for neuroscience in the court, likening it to *Brown v. Board of Education of Topeka*, the landmark 1954 ruling that ended racial segregation in public schools. *Brown v. Board* was believed to be heavily influenced by a sociological study of children judging dolls of different colours. This purportedly showed that segregation had a negative impact on black students' self-esteem. Just as that case is thought to mark the modern era of the court's use of scientific research, *Roper v. Simmons* was believed to herald the era of neuroscience in the court.

Emboldened by the *Roper* decision, advocates and attorneys have increasingly called on neuroscience research when arguing that juveniles should be spared the harshest punishments. In 2010, the court ruled in *Graham v. Florida* that sentencing a juvenile to life without parole for a crime other than homicide was "cruel and unusual". The neuroscience

research, emphasized in briefs from the AMA, the APA and Graham's lawyers, seemed to strike even more of a chord with the justices, who wrote in the majority opinion that "developments in psychology and brain science continue to show fundamental differences between juvenile and adult minds".

BRAIN OVERCLAIM

But some researchers are uncomfortable with the way this research is being used in the criminal justice system. There is certainly a correlation between brain development and behavioural maturity, but saying that observed differences in the adolescent brain cause certain adolescent behaviours is a symptom of what lawyer and psychologist Stephen J. Morse at the University of Pennsylvania in Philadelphia calls "brain overclaim syndrome". Neuroscientists, says Morse, are "always using language that suggests causation, when they don't know causation".

Moreover, the behaviours studied in the lab bear little resemblance to criminal behaviour. Scientists cannot perform brain imaging on someone in the process of committing a crime. Instead, they study volunteers who are challenged to resist impulses or follow rules. In some experiments looking to assess impulse control, for example, volunteers are told not to look at a flashing light. Not looking at the light, says Luna, who has used the task for more than ten years, "is extremely difficult to do. We know exactly what brain circuitry gets engaged to stop that reflexive response. Is it the same as murdering someone? No." But Luna argues that such experiments are relevant. Although adolescents can stop themselves from looking at the light, doing so is associated with much more activation of their prefrontal cortex than is seen in adults, suggesting that it is more difficult for them to control their impulses⁷.

But that wasn't really in question. Decades' worth of behavioural research had already established that teenagers take greater risks, are more driven by emotions and peer influence, and generally behave less responsibly than adults. Still, advocates say, neuroscience bolsters the arguments. "When you come up with a biological mechanism, it gives it a lot more credibility," says Luna. To many audiences, neuroscience evidence tends to be more convincing than behavioural science, particularly when aided by pictures of brain areas lighting up.

"Lawyers and judges have grown up thinking that social science is soft," says constitutional law scholar David Faigman at the University of California Hastings College of the Law in San Francisco. "Neuroscience gives the courts a hook." But, Morse says, treating neuroscience data as somehow harder than behavioural data is logically flawed. The neuroscience is

► NATURE.COM
Read more about
science in the
courtroom at:
go.nature.com/ez6pwk

only relevant to the law when it connects with behaviour. "If the psychology is soft, the neuroscience is soft," Morse says.

Hard or soft, the data are open to interpretation. And scientists have little control over when or how their research is used as a persuasive tool. Steinberg's research was cited in the court's opinion in *Roper v. Simmons* and in briefs in support of juveniles for *Graham v. Florida* and the two current cases — but it has also been used by advocates seeking to restrict the rights

JUDGES MAY USE SCIENCE FOR RHETORICAL LEVERAGE, BUT SCIENCE RARELY, IF EVER, DRIVES A COURT DECISION.

of adolescents, in particular the right to have an abortion without parental permission.

Ultimately, the neuroscience data can only take legal arguments so far. Juveniles may be less responsible than adults, but they may still be responsible enough to deserve the same punishment for similar crimes, says Morse. "That's not a scientific question. It's a moral question, and, ultimately, a legal question." Because of this incompatibility, Faigman says that science rarely, if ever, drives a court decision. Judges may use science for rhetorical leverage when writing the opinion, but "it's not always clear whether the research made an impact".

HOT HEADS OR COLD BLOODED

Take *Roper v. Simmons*. Briefs to the court supporting Simmons presented scientific evidence that the immature brains of adolescents like him render them more likely to act in the heat of the moment, giving little thought to the future. But Simmons didn't fit that description: according to Supreme Court documents, he told his friends that he "wanted to murder someone" days before the murder, "by breaking and entering, tying up a victim, and throwing the victim off a bridge".

Although the court alluded to scientific research in its opinion, that research may not have influenced the decision. "The justices may have [already] come to the conclusion that the death penalty was fundamentally flawed, at least in its application to juveniles," says Faigman. "They needed to support that with good reasons."

If science's part in *Roper v. Simmons* was

merely as a makeweight, then comparing it to *Brown v. Board* might not be far off the mark. In that case, Faigman says, the court had probably already decided to end segregation in public schools on moral and social grounds. Although "the perception was that science lay at the core of the decision", says Faigman, the doll study had been "severely criticized" before the ruling. It was only mentioned in a footnote.

When the court decides on *Miller v. Alabama* and *Jackson v. Hobbs*, advocates for juveniles hope to see life without parole taken off the table for all juveniles. But the court may draw the line at 14, the age of the petitioners. Whatever line the justices do draw, if they draw one at all, it will probably be based on legal and moral ideas rather than scientific evidence.

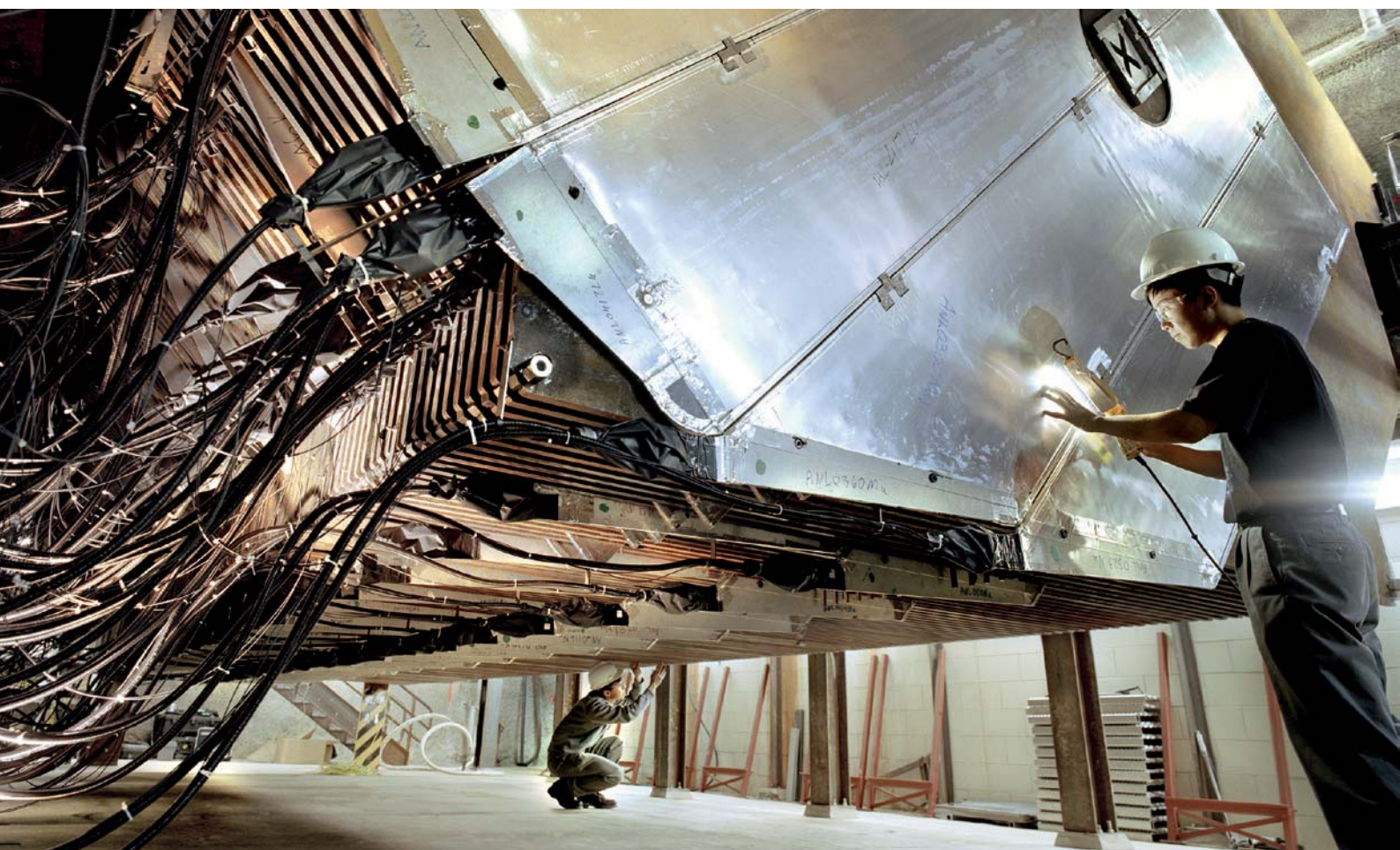
So does neuroscience deserve a place at the table? Perhaps, says Terry Maroney of Vanderbilt University Law School in Nashville, Tennessee, but it shouldn't outshine other evidence. Solid research by social scientists shows, for example, the high potential of teen criminals to reform, the inability of life sentences to deter juvenile crime, and racial inequalities in life-without-parole sentencing⁸. Focusing on biology draws attention away from socioeconomic, educational and cultural drivers of criminal behaviour. After all, most teenagers do not commit murder, despite their fledgling frontal cortices.

Evan Miller grew up with physical and emotional abuse so severe that he attempted suicide several times, starting at the age of five. He was using drugs and alcohol regularly before he was a teenager; the man he killed was his mother's drug dealer. Kuntrell Jackson was raised in the lap of violence. Bryan Stevenson, his attorney, told the Supreme Court justices that "his grandmother shot his uncle. His mother shot a neighbour. His brother shot someone. They were all put to jail."

Neuroscience supports what centuries of casual observation have strongly suggested: teenagers tend to be more impulsive and irresponsible than adults. That probably can't be changed, but the environments that they grow up in can be addressed with social policy. "That's harder," says Maroney. "It requires more work and sacrifice. It's easier to look and say, 'Ooh, look what's going on inside their heads! It's all their problem.'" ■

Lizzie Buchen is a freelance reporter in San Francisco and volunteers with prisoners at San Quentin State Prison.

- Liston, C. *et al.* *Cereb. Cortex* **16**, 553–560 (2006).
- Velanova, K., Wheeler, M. E. & Luna, B. *Cereb. Cortex* **18**, 2505–2522 (2008).
- Olson, E. A. *et al.* *J. Cogn. Neurosci.* **21**, 1406–1421 (2009).
- Burnett, S., Bird, G., Moll, J., Frith, C. & Blakemore, S.-J. *J. Cogn. Neurosci.* **21**, 1736–1750 (2009).
- Hare, T. A. *et al.* *Biol. Psychiatry* **63**, 927–934 (2008).
- Galvan, A. *et al.* *J. Neurosci.* **26**, 6885–6892 (2006).
- Luna, B. *et al.* *NeuroImage* **13**, 786–793 (2001).
- Nellis, A. *The Lives of Juvenile Lifers* (The Sentencing Project, 2012).



Two scientists work on the near detector of the MINOS neutrino experiment at Fermilab in Illinois.

A MATTER OF DETAIL

AMERICAN NEUTRINO PHYSICISTS ARE GETTING THE MEASURE OF THEIR QUARRY IN ULTRA-HIGH PRECISION.

EUGENIE SAMUEL REICH

“It’s like Christmas shopping at the specialist boutiques,” says Phil Adamson, as he describes his recent US\$250,000 buying spree. Adamson, a physicist at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, leads a team that has spent the past few months acquiring and installing three ultra-high-precision atomic clocks; six Global Positioning System receivers; more than a kilometre of optic fibre; two auxiliary detectors and at least one pair of timing-interval counters (“kind of fancy stopclocks”, he says) — all to time subatomic neutrinos with nanosecond precision as they pass through the detectors of Fermilab’s Main Injector Neutrino Oscillation Search (MINOS).

The researchers are trying to answer one simple question: do neutrinos travel at or below the speed of light, as required by the theory of relativity, one of the most fundamental tenets of modern physics? Or do they travel just a tiny fraction faster, as suggested with enormous fanfare last September¹ by an experiment in Italy?

To outsiders, this debate has already been settled. Researchers on the Italian experiment, based at Gran Sasso National Laboratory near L’Aquila, announced in March that they had found the error in their measurement — and two of the team’s leaders resigned (see go.nature.com/xjzhqa). At the same time, physicists working at a different detector at Gran Sasso have published measurements² showing that neutrinos do indeed obey the light-speed limit.

Yet neutrino speed is still a prime focus for the MINOS physicists, who are carrying out their own high-precision measurement — not least because they did not do this when they saw hints of a faster-than-light neutrino in their own data some years ago. “This is so much on our front burner,” says Robert Plunkett at Fermilab, a spokesman for MINOS. Around 30 members of the 150-strong collaboration are now working on the search.

P. GINTER

Precision is what the MINOS team is all about. In 2008, Fermilab's Tevatron was supplanted as the world's highest-energy particle accelerator by the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory outside Geneva, Switzerland. Since then, US particle physicists have moved from studying collisions at the highest energies to working with beams at the highest intensities, adapting the country's existing accelerator facilities to measure the rates of extremely rare interactions. The hope is that, even at lower energies, forcing theory and experiment into close comparison can turn up anomalies that point to new physics.

Neutrino physics is the centrepiece of that programme. The particles respond only to the aptly named weak force, and mostly stream through solid matter as if there were no barrier. But this aloofness also makes neutrinos potentially a very clean probe of exotic forces: when a neutrino does hit another particle, physicists don't have to disentangle the effects of the much larger strong and electromagnetic forces. Experiments using accelerators, in which physicists can control the energy and direction of the neutrino beam, take maximum advantage of that fact. And MINOS is the most sophisticated experiment of this kind in the United States.

True, it takes a special kind of researcher to pursue such work — pushing measurements towards the last possible decimal point, over decades if need be. But most of the MINOS physicists share the attitude voiced by team member Nathaniel Tagg, a physicist at Otterbein College in Westerville, Ohio. "I'm one for long shots."

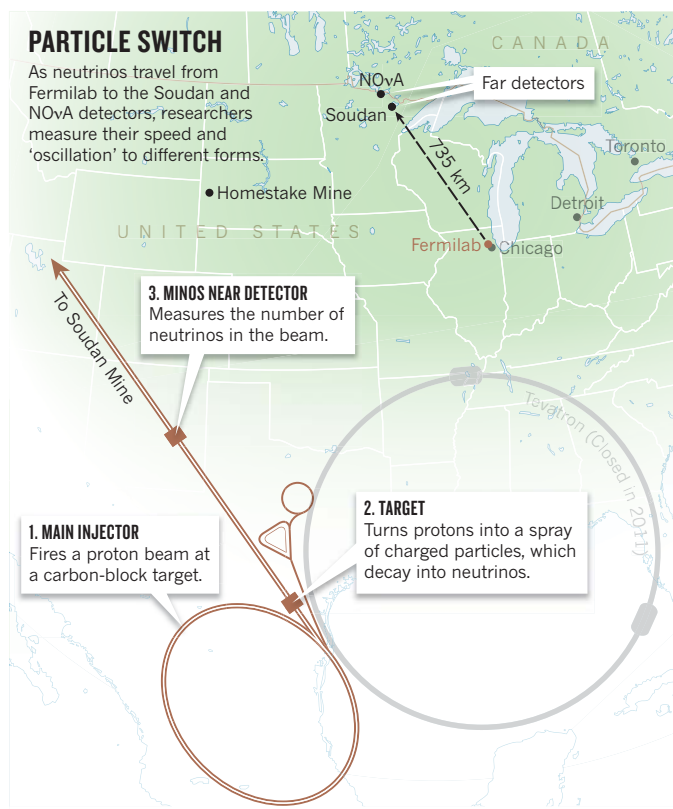
COWBOY PHYSICS

Neutrinos' central role in US physics research is a case of historical turnabout, says Tagg. For years the field was considered 'cowboy physics' — a fringe area best left to diehard experimentalists who were willing to drag their detectors down remote mineshafts and into underground bunkers, where the devices could escape the confounding effects of atmospheric cosmic rays.

The archetypal cowboy was Raymond Davis, a radiochemist at Brookhaven National Laboratory in New York, who wanted to make the first direct observations of nuclear fusion reactions in the core of the Sun. He began in 1967 by setting up a 380-cubic-metre tank of perchloroethylene, a dry-cleaning fluid, nearly 1,500 metres down in the Homestake gold mine in Lead, South Dakota. Davis's idea was that neutrinos created in the Sun's nuclear reactions would very occasionally strike a chlorine atom in the tank and turn it into a radioactive isotope of argon, which he could extract and detect by chemical means. But only about one-third of the expected number of neutrinos seemed to be showing up. Either the theorists were wrong about the rate of fusion reactions in the core of the Sun — which didn't seem likely, given the success of their calculations in other areas of astrophysics — or two-thirds of the neutrinos were getting lost.

It took decades for physicists to reach a consensus that this 'solar neutrino problem' was real, not some obscure experimental error, and only in the 1990s did they converge on a probable solution. This started from the hypothesis that neutrinos come in three types, or 'flavours', each of which is the electrically neutral partner of a negatively charged particle with mass — an electron, a muon or a tau. The Sun's fusion reactions produce only electron neutrinos. But as soon as they are produced, according to the theory, these neutrinos begin to 'oscillate', changing from one flavour to another as they travel. By the time they get to Earth, the three flavours have mixed themselves into equal proportions — meaning that

NATURE.COM
For more on the
OPERA neutrino
experiment, see:
go.nature.com/eswpxs



just one-third will be electron neutrinos, the only kind that Davis's tank could detect.

This oscillation theory elegantly accounted for the missing solar neutrinos, but posed a new conundrum. The standard model of particle physics held that all three flavours of neutrino have a mass of exactly zero, like that of the photon. But oscillation was possible only if neutrinos have at least a very small mass — which meant that the particles were hinting at some kind of physics beyond the standard model.

This prospect quickly took neutrinos out of the cowboy-physics category (Davis himself would share the Nobel Prize in Physics in 2002). MINOS was one result (see 'Particle switch'). Planning began in the 1990s as physicists looked for a way to verify the oscillation model. In particular, they wanted to create an artificial neutrino beam that would remove one of the biggest uncertainties in solar neutrino experiments, which was the number of neutrinos arriving at the detector. The plan called for Fermilab's accelerator complex to fire protons at a graphite target, producing a spray of short-lived charged particles that would decay into muon neutrinos. The resulting neutrino beam would pass through a 1,000-tonne, train-carriage-sized 'near' detector in the Fermilab grounds, where enough neutrinos would be captured to gauge the total number in the beam. The remaining particles would continue underground on a 735-kilometre, straight-line path to the Soudan Mine in northern Minnesota, where a five-times-larger 'far' detector would measure how many muon neutrinos arrived. If some had oscillated into other forms, the discrepancy would be obvious — and neutrino oscillation would be verified.

COLLISION COURSE

In 1998, however, while MINOS was still on the drawing board, the US project was beaten to this goal by the Super-Kamiokande experiment near Hida in Japan. Through a clever experimental design, the Japanese physicists had been able to verify the existence of oscillation using neutrinos

generated by cosmic rays striking the atmosphere³.

That disappointment took a toll, admits Adamson, one of many particle physicists drawn to the field by a desire to sort out the oscillation phenomenon. But it also proved an opportunity. Rather than being the ones to discover new physics, the MINOS team decided that they would be the ones to carefully characterize the phenomena that the Super-Kamiokande had found. "By the time we started to operate [in 2005] we were trying to make precision measurements," says Adamson.

One obvious question was why neutrinos have mass at all. The standard model can explain the mass of charged particles, such as the electron, as a subtle interaction with the hypothetical Higgs boson. But that mechanism was not supposed to affect neutrinos. The three neutrino varieties have masses so tiny, less than one-millionth that of the electron, that some kind of exotic mass-generating mechanism may be at work, and high-accuracy measurements of the oscillation phenomenon could shed light on what that is.

Such questions soon began to move the long-term planning for accelerator neutrino facilities to the forefront of US high-energy physics. In 2006, the Particle Physics Project Prioritization Panel of the Department of Energy (DOE) laid out a roadmap for the field that included continued support for MINOS and endorsement of a new experiment known as the NuMI Off-Axis Neutrino Appearance (NOvA). (NuMI, which stands for 'neutrinos at the main injector', is the name of the neutrino beam serving the experiment.)

ACCELERATED DEVELOPMENT

The plans for NOvA called for a boost in the energy of Fermilab's neutrino beam and a new detector farther north in Minnesota. One of its key goals would be to measure how oscillation occurs among neutrinos' three antimatter counterparts, the antineutrinos, and find out whether the process obeys 'charge-parity' symmetry. This symmetry, which basically means that interactions should remain unchanged if particles and antiparticles swap places and everything is viewed in a mirror, is known to be violated in only a few, very rare reactions. Nevertheless, charge-parity violation is thought to be the ultimate explanation for the emergence of much more matter than antimatter from the early Universe, and why stars, planets and living things can exist today. If the symmetry can be violated for neutrinos and antineutrinos, then these ghostly particles could provide unique insights into the processes that have made the Universe the way it is.

At MINOS, meanwhile, the team was busy boosting the intensity of Fermilab's neutrino beam, setting a world record for precisely determining the difference between the masses of the three types of neutrino⁴, and measuring a variety of parameters crucial for the design of NOvA.

Then, in June 2010, scientists at MINOS reported early signs of a discrepancy in the rates at which neutrinos and antineutrinos oscillated⁵. This particular discrepancy would have violated another fundamental symmetry of quantum field theory known as CPT, for charge, parity and time. In its way, this would have been just as astonishing as faster-than-light neutrinos. MINOS scientists began to hope for a paradigm-shattering discovery. But these hopes were dashed when further data, reported in February this year⁶, suggested that the result was a statistical fluctuation. "It was disappointing," says Justin Evans, a physicist at University College London who is a member of the MINOS team. "We made the world's most precise measurement of antineutrinos' parameters — but everyone wants to be the group that discovers something new."

The faster-than-light neutrino announcement last September from the Oscillation Project with Emulsion-tracking

Apparatus (OPERA) experiment in Gran Sasso shook the MINOS team out of its data-collecting routine. The Italian lab reported that neutrinos seemed to be making the 730-kilometre trip from CERN to Gran Sasso some 60 nanoseconds faster than a light beam would. The announcement galled some MINOS collaborators, who back in 2001 had proposed an ultraprecise measurement of neutrino speed only to have the idea nixed by the DOE. The team decided not to resurrect the idea in 2007, when its low-precision measurements hinted that neutrinos might be travelling faster than light⁷. It didn't seem worth fighting that battle again to follow up a result with minimal statistical significance.

So, stung by the hullabaloo from Italy — and aware that MINOS, given the similarity of its set-up to OPERA, was uniquely placed to provide the all-important independent replication of the remarkable finding — the MINOS team was determined to get the measurement right.

"IT'S A PLEASURE AND PRIVILEGE TO BE IN A POSITION TO SETTLE THIS."

But MINOS was scheduled for a temporary shutdown in March 2012, during which the energy of its neutrino beam would be boosted to serve NOvA. MINOS would also receive some upgrades so that it could look for exotic phenomena such as sterile neutrinos — hypothetical particles that would not participate in any interaction governed by the standard model — and neutrino oscillation into extra dimensions. Yet everyone understood how momentous it would be if neutrinos really did violate Einstein's speed limit, says Adamson. Once the team realized that it had the people and the set-up to check the result in a reasonable time, he says, it felt it had a duty to weigh in.

Fermilab approved a two-month delay in the neutrino beam shutdown, and the MINOS team resurrected and improved its decade-old proposal to do the measurement at high precision. By the time of the shutdown, now scheduled for 1 May, the experiment should have yielded data sufficient to pinpoint the time-of-flight measurement to within 11 nanoseconds, similar to OPERA's uncertainty. And by sometime in 2013, after the neutrino beam is upgraded to higher energy, the increase in event rate should allow MINOS to reach an error of between 2 and 7 nanoseconds.

Physicists at MINOS are determined to make the measurements, even if their chance of a faster-than-light finding is now exceedingly remote given OPERA's admitted error. "There's a chance it will turn out to be really interesting," says Tagg. But if, as looks likely, the measurement simply confirms something that everybody already knows, MINOS will at least have produced a high-precision measurement of a fundamental parameter, something that physicists on the experiment insist is a noble contribution, even if it's not a discovery. "It's a pleasure and privilege to be in a position to settle this," says Plunkett. ■

Eugenie Samuel Reich is a contributing correspondent for Nature.

1. Adam, T. *et al.* Preprint at <http://arxiv.org/abs/1109.4897> (2011).
2. Antonello, M. *et al.* Preprint at <http://arxiv.org/abs/1203.3433> (2012).
3. Fukuda, Y. *et al.* *Phys. Lett. B* **436**, 33–41 (1998).
4. Adamson, P. *et al.* *Phys. Rev. Lett.* **106**, 181801 (2011).
5. Adamson, P. *et al.* *Phys. Rev. Lett.* **107**, 021801 (2011).
6. Adamson, P. *et al.* Preprint at <http://arxiv.org/abs/1202.2772> (2012).
7. Adamson, P. *et al.* *Phys. Rev. D* **76**, 072005 (2007).

COMMENT

EXHIBITION London show reveals Leonardo's anatomical sketches **p.314**

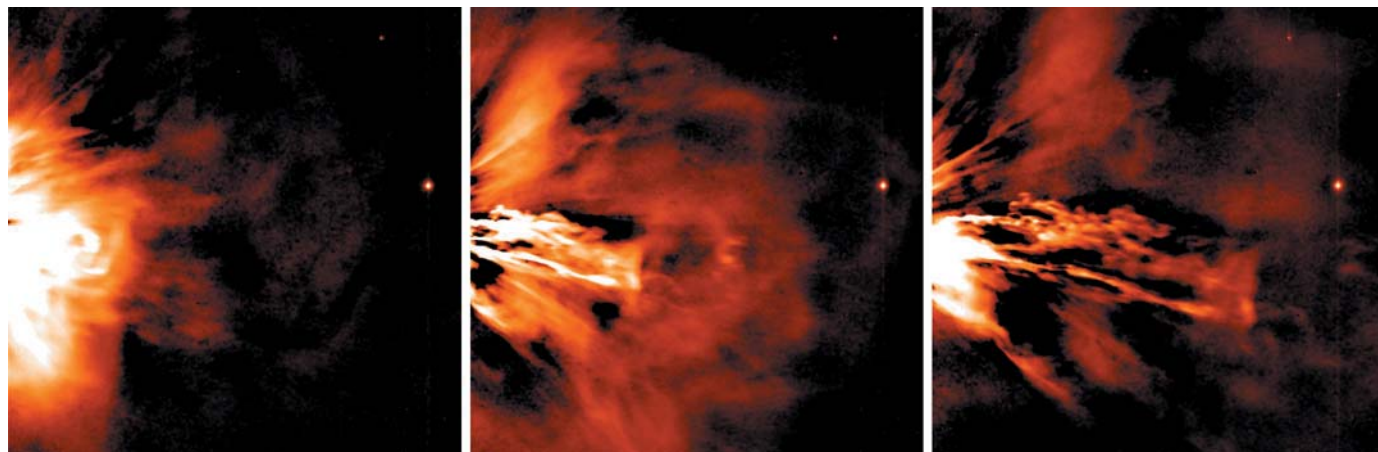


GENDER Evolutionary reasons why women may stall on the career ladder **p.317**

GENOMICS Reward companies for investing in personalized medicine **p.318**

AGRICULTURE Adding micro-nutrients to soil could boost crop yields in Africa **p.318**

J. DAVIES, STFC/RAL SPACE



This eruption of plasma from the Sun in June 2011, captured by instruments on the STEREO spacecraft, didn't cause a space weather storm on Earth. Others will.

Prepare for the coming space weather storm

We need to improve estimates of geomagnetic storm size, says **Mike Hapgood**, so we can be ready for huge disruptions to electrical systems.

On 13–14 March 1989, Earth experienced its largest geomagnetic storm in decades: a barrage of hot, ionized particles streamed towards the planet, wreaking havoc on its magnetic field and electrical systems. The storm caused a power blackout in Quebec, Canada, leaving 5 million people without electricity for 9 hours in cold weather and causing Can\$2 billion (US\$2 billion) in damages and losses to business. It also permanently wiped out a US\$12-million transformer in the United States, and caused two large UK transformers to be sent back to the factory for repairs. Space agencies temporarily lost track of some 1,600 spacecraft.

We should be prepared for much worse. Even bigger storms have been recorded in the past, although society at the time was less reliant on electrical systems and thus less vulnerable. In September 1859, a huge geomagnetic event disrupted the then-emerging technology of the electric telegraph, shocking

operators and sparking fires in telegraph offices. A similar event in May 1921 burnt down a telephone exchange at Karlstad in Sweden. A storm on these scales today would prove disastrous to our highly electronic society. A study by the UK National Grid suggests that a repeat of the 1859 event would leave some regions without power for several months¹. In the United States, some assessments predict wide-scale disruption with ripple effects lasting for years and an economic impact of several trillion dollars².

The source of all this damage is coronal mass ejections (CMEs) — huge eruptions of magnetically charged plasma that occur during magnetic storms in the Sun's atmosphere and that increase the stream of particles in the solar wind by many hundreds of times. When CMEs reach Earth, they can inject large amounts of energy into the planet's magnetic field, causing field variations and driving extra current into electricity grids. The storms can heat and expand the upper

atmosphere, nudging the orbits of satellites and space debris, and can charge spacecraft components to dangerous voltages.

There is some capability for forecasting space weather in the short term. The US Space Weather Prediction Center in Boulder, Colorado, can provide warning for strong geomagnetic storms 10–60 minutes in advance with better than 50% reliability, using observations of approaching CMEs by NASA's Advanced Composition Explorer (ACE) spacecraft. That provides a small window in which to apply countermeasures to protect large systems such as the power grid. The newer NASA Solar Terrestrial Relations Observatory (STEREO) has shown that better tracking of CMEs might give reliable warnings 6 hours in advance, with one-hour accuracy. The need to keep improving forecasts was shown in March 2012, when a major CME approaching Earth sparked intense media coverage. Forecasts the day before predicted arrival times that varied ▶

► by up to 18 hours. Many were inaccurate, although an experimental forecast from the National Oceanic and Atmospheric Administration (NOAA) in Boulder was close: just one hour early. In the end, that CME produced only a modest space weather event, thanks to the direction of the impacting magnetic fields, which cannot yet be forecast.

In the long term, we still have little sense of what maximum space weather event we should prepare for. Many at-risk electrical systems are designed to withstand events of the type seen in the past 40 years or so of engineering: many power grids, for example, now require that new transformers are able to withstand conditions similar to those experienced in 1989. Last year's earthquake and tsunami in Japan show the dangers of preparing only for an event similar to that seen in recent decades. Instead, we should prepare for a space weather event that might happen only once in 1,000 years.

To do this, scientists need to improve the availability of older space weather data, develop more-sophisticated models to predict future scenarios and ensure that the right systems — from the electric grid to the Global Positioning System (GPS) time-stamps relied on for financial transactions — are prepared properly. That shift is happening, but not quickly enough.

PREPARING FOR THE WORST

The media often airs concerns about solar activity — there have been many stories about the handful of giant solar flares that have occurred since January, for example. But solar flares have a minimal effect on space weather; although they are a good indicator of the amount of energy being released in a solar event, they do not cause CMEs and are not indicators of geomagnetic storms. The media also tends to note that the Sun is heading towards a maximum in its 11-year activity cycle. But the Sun's activity is more complicated than this. Superimposed on the 11-year cycle is a much longer cycle that has produced a series of 24 'grand maxima' over the past 9,000 years. The latest grand maximum, which started in about 1920, is now coming to an end, so the next solar maximum, due in 2013 or 2014, should be unusually low (see 'Solar cycles').

These facts should not cause complacency, however. Although a solar maximum might bring a higher frequency of more intense storms on average, there is no evidence that these long-term trends affect the intensity of any individual burst of solar activity — indeed, the 1859 event occurred outside of a grand maximum. We need to develop safeguards against the entire range of possible events that can be generated by CMEs.

The first step towards this goal involves increasing the availability of data, so that statistical methods can be used to improve

estimates of worst-case scenarios. A 2011 statistical study³ of digital magnetometer data taken in northwest Europe since 1980, for example, looked at the intensity of an event that is likely to occur once every 200 years. For the short-time-frame magnetic field variations (in the order of minutes) that threaten power grids, the study estimated a worst-case intensity in the range of 1,000–6,000 nanoteslas per minute. Fortunately, this roughly matches the US and UK estimates of a worst case of 5,000 nanoteslas per minute, which was taken from one observation from central Sweden during the May 1921 event.

Most historical data sets exist only on paper as charts or tables, sometimes handwritten. These include ionospheric data going back

"Geomagnetic storms pose a serious threat to our technology-dependent society."

80 years and magnetic data going back 170 years. Over the past decade, much progress has been made in identifying these historical records; they now need to be made usable through

digitization. The advent of web-based citizen-science projects may offer a way forward, with volunteers transcribing printed or handwritten values, or converting graphs to tables. This approach is already being used to analyse handwritten historical temperature records for climate-change studies. There is a large public appetite for this type of engagement, and such work can prove scientifically useful for space weather forecasting. In the Solar Stormwatch project, for example, volunteers measure features of CMEs in real time, and the measurements are then used to track and forecast the arrival of the eruptions at Earth.

A second important step to safeguarding against CMEs is developing physics-based modelling: models that are based on a fundamental understanding of CMEs, rather than relying solely on what has been seen before.

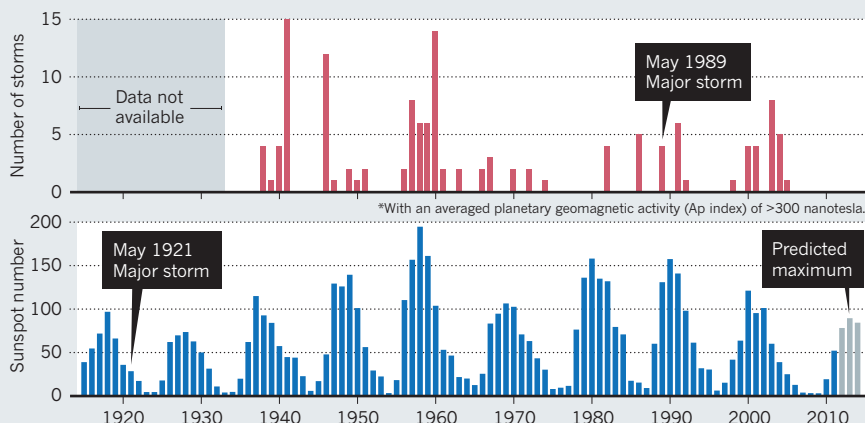
These models require an understanding of how CMEs travel through interplanetary space and inject energy into Earth's magnetosphere, and of how the resulting geomagnetic storm alters the composition, temperatures and velocities of the upper atmosphere. In this way, extreme events can be simulated before they happen. However, such modelling is still in its infancy, comparable with the situation in atmospheric modelling 50 years ago. Most efforts today focus on specific elements in the chain. A partnership established in 2011 between the NOAA and the UK Met Office — part of increasing UK-US collaboration on space weather — is aiming to build a more integrated, practical model for forecasting these events.

Models for space weather forecasts need to include long-term changes to Earth: its magnetic field has declined by 10% since 1850, and the height of upper atmospheric layers may be changing in response to climate change. Physicists need to learn how these factors will affect space weather, and that can be done only through good physics-based modelling.

The models are also limited by the complexity of some of the physics involved. For example, extreme space weather conditions create a strong flow of plasma of perhaps many kilometres per second in the ionosphere. For computational efficiency, models generally treat this plasma as a magnetized fluid in thermodynamic equilibrium. This simple description works well when modelling large-scale features, but it misses the rich set of plasma phenomena that arise during strong plasma flows. In these very energetic conditions, modellers should expect significant deviations from thermodynamic equilibrium. Physicists need to develop small-scale models that capture this full range of plasma physics, and then find computationally efficient ways to incorporate them into large-scale simulations.

SOLAR CYCLES

The number of severe* space weather storms each year (top) is related to, but not precisely determined by, short- and long-term patterns in solar activity (bottom).



SOURCE: SWPC/NOAA



Space weather, which sometimes causes auroras on Earth, threatens to play havoc with electrical grids and to disrupt signals from the Global Positioning System.

A third, more speculative, way to safeguard against CMEs is to look at space weather around other stars. We can observe flares on other stars, for example, using X-ray space telescopes such as the European Space Agency's X-ray Multi-Mirror (XMM)-Newton. In the near future, we may also be able to observe radio and near-infrared emissions from auroras on gas-giant planets around those stars. That will give us insight into the power carried by their stellar winds, in particular by CMEs. To make a valid comparison with our own space weather, we will need to focus on Sun-like stars with similar rotation rates to the Sun. At present, this limits such studies to something between two and ten discovered stars. The search for such stars is already a hot topic in stellar astronomy, and should be encouraged as an area that has the potential to broaden our understanding of space weather.

FUNCTIONAL KNOWLEDGE

Alongside improvements to our knowledge of worst-case-scenarios, we must also ensure that crucial industries are prepared. The electrical power industry has a good awareness of the geomagnetic threat after the 1989 failure in Quebec. Power grids can and do harden their systems by installing transformers that are more resilient to power fluctuations, and by having enough reserve generators to weather a storm.

In aviation, severe space weather can black out communications links to the ground, disrupt on-board electronics and increase the radiation dose accumulated by aircrew and passengers. There is some awareness of these risks: airlines operating transpolar routes divert to longer routes at lower latitudes during bad space weather (incurring costs of

extra fuel and crew hours). The US Federal Aviation Administration is supporting work to establish international standards for space weather forecasts for preflight briefings.

Further work will be needed for both industries as the worst-case scenarios become clearer. For example, the power sector will need to decide whether it is cost-effective to add expensive devices to power grids to block the extra currents generated by space weather, as has been done in Quebec. Right now, such discussions are not based on sufficient facts.

A less obvious risk is the potential loss of the accurate timing signals provided by GPS. Many industries exploit these signals: for example, the finance industry needs to timestamp transactions with millisecond accuracy to support automated trading. This threat is easily overcome by using high-precision local clocks that require only occasional synchronization with GPS, perhaps once every few weeks. In this way, industries could survive a severe space weather event lasting one or two days. When I was preparing a report on solar storm impacts with Lloyd's insurance in 2010, it was unclear whether industries that rely on accurate timing are taking such steps. Private operators should be encouraged to be open about their efforts to safeguard against space weather.

National organizations that assess space weather risks include the UK Civil Contingencies Secretariat and the US Federal Emergency Management Agency. Several international bodies also have growing space weather efforts, including the World Meteorological Organization in Geneva, Switzerland, and the European Commission, the latter of which held a Space

Weather Awareness Dialogue in October 2011. We should combine these efforts through an international network, bringing together all aspects of risk and implication assessments, operational responses and policy development.

This area of science has moved away from its roots in astronomy and communication engineering and is now better handled as a generic environmental risk to society and the economy, in parallel with earthquakes, volcanoes and floods. National and international funding organizations are increasingly transferring relevant funds to environmental research bodies. This is a welcome development that should be encouraged.

Geomagnetic storms pose a serious threat to our highly vulnerable, technology-dependent society. We need a much better understanding of the likelihood of space weather disruptions and their impacts, and we need to develop that knowledge quickly. ■

Mike Hapgood is head of the space environment group with RAL Space, part of the UK Science and Technology Facilities Council. He is based at the Rutherford Appleton Laboratory (RAL) in Didcot, UK, and chairs an expert group that advises the UK government on space weather risks. e-mail: mike.hapgood@stfc.ac.uk

1. *Developing Threats: Electro-Magnetic Pulses (EMP)* pp. Ev26–Ev27 (UK House of Commons Defence Committee, 2012); available at <http://go.nature.com/edi8qf>
2. National Research Council Space Studies Board *Severe Space Weather Events — Understanding Societal and Economic Impacts* (National Academies Press, 2008); available at <http://go.nature.com/gccflj>
3. Thomson, A. W. P., Dawson, E. B. & Reay, S. J. *Space Weather* **9**, S10001 (2011).

MEDICINE

Leonardo's anatomy years

A London exhibition will expose the Renaissance master's staggering medical discoveries, which languished unpublished for centuries, explains **Martin Clayton**.

Leonardo da Vinci is the archetype of the Renaissance man, but since his day he has been seen primarily as a painter who dabbled in the sciences. Leonardo would not have recognized this image: his scientific studies were as important to him as his art. Of all his investigations — which included optics, geology, botany and hydrodynamics — the field that engaged him most was human anatomy.

Eighty-seven of Leonardo's finest sheets of anatomical study from the UK Royal Collection's total of more than 550 by the artist are to go on show from May at the Queen's Gallery in London. They range from a fetus in the uterus to the vertebral column, and will constitute the largest-ever exhibition of Leonardo's anatomical work, and the first in the United Kingdom since 1977.

The exhibition represents two campaigns of intense work around 1490 and between 1507 and 1513, during which Leonardo dissected around 30 human corpses. He fully intended to publish his findings, and had he done so, he would have transformed the study of anatomy in Europe. Many of the bodily structures that he depicted would not be described again for centuries. But Leonardo's perfectionism, his difficulties in reconciling his observations with established beliefs about physiology, and simple bad luck prevented him from bringing his work to a conclusion.

We know little about Leonardo's early life. The illegitimate son of a notary and a peasant girl, he was born in 1452 near the town of Vinci in Tuscany, Italy. He learned to read and write but his arithmetical skills were shaky, and although he taught himself Latin as an adult, he was never comfortable with the language of most contemporary scientific writings. He was to portray this as an advantage — claiming that he was a “disciple of experience” unencumbered by ancient belief — but in reality his position outside the scientific-philosophical mainstream would deny him the range of contacts and resources that his peers enjoyed.

In his twenties, Leonardo joined the painters' guild in Florence and worked in the

studio of the innovative sculptor and painter Andrea del Verrocchio, where he picked up the rudiments of engineering. Then, in the 1480s, Leonardo moved to Milan, where his range of interests widened at a remarkable rate and he began to compile material for a treatise on the theory of painting.

of the body: Leonardo also wanted to learn how an individual's appearance from moment to moment was related to the workings of the mind,

Leonardo da Vinci: Anatomist
The Queen's Gallery,
London. 4 May –
7 October 2012.
www.royalcollection.org.uk

so that a painting would reveal the emotions of the protagonists and the human drama of the scene.

Leonardo thus aimed to understand the perception of reality through the senses, the structure of the mental faculties and how the nerves configure the muscles and bones. How could he even begin to investigate these topics? Human dissection was not banned, as is often supposed; indeed, a papal bull of 1482 expressly permitted it. But Leonardo was a mere craftsman, and — then as now — a craftsman could not simply acquire a corpse and start cutting it up. Instead, he was reliant at first on animal dissection, traditional belief and simple speculation.

In 1489, Leonardo obtained a human skull. He sectioned it to investigate its internal structure, recording his findings in exquisite drawings in a notebook annotated in his habitual mirror writing (Fig. 1). But the shape of the skull (and by implication of the brain) failed to provide Leonardo with any useful information about the relationship between mind and body. The gulf between Leonardo's ambitions and his achievements brought his first wave of anatomical research to a halt. In his

mural *The Last Supper*, finished in the 1490s, the disciples' poses and expressions convincingly capture their varied emotions, but owe nothing to his anatomical investigations.

Leonardo returned to the subject years later, in the wake of a commission to paint a huge battle scene in the Palazzo della Signoria in Florence. In preparation for the project he executed many drawings of male musculature, some so detailed that they can only have been based on flayed bodies. Leonardo was then in his fifties and one of the most ▶

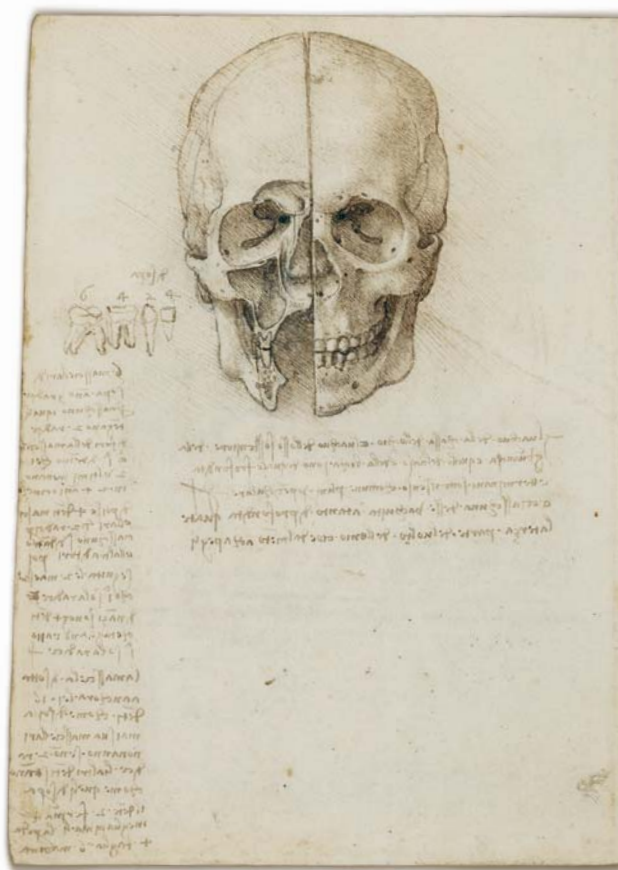


Figure 1 | *The skull sectioned* (1489), one of Leonardo da Vinci's earliest anatomical studies, did not yield the insights that he craved.

Leonardo had come to see painting as a scientific activity, in which every effect (light and shade, colour, perspective) and form should be based on a true understanding of nature. The human body was the principal subject of the Renaissance artist and Leonardo

soon realized that he would have to devote a separate treatise to it. It was not sufficient to study the permanent anatomy

➔ **NATURE.COM**
For a video about
Leonardo's
anatomical studies:
go.nature.com/cxnnqn

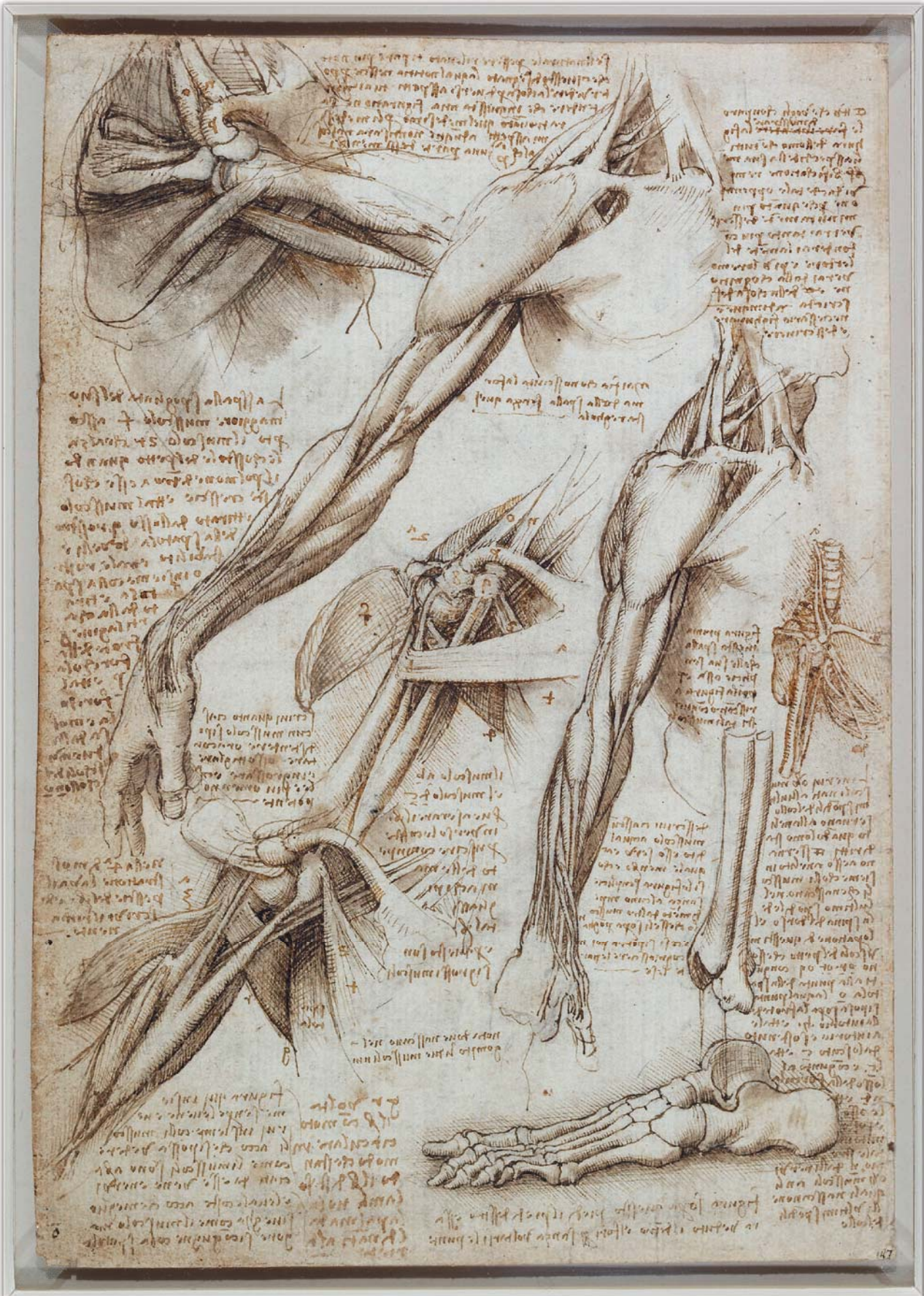


Figure 2 | With *The muscles of the shoulder and arm, and the bones of the foot* (c.1510–11), Leonardo sought to explore the dynamic interplay of bodily elements that underlay the poses of human models in his paintings .

► celebrated artists in Italy; at this point in his career, he seems to have had little difficulty in obtaining permission to dissect human corpses.

In the winter of 1507–08, Leonardo witnessed the peaceful demise of an old man in a hospital in Florence, and wrote in his notebook that he performed a dissection “to see the cause of so sweet a death”. He attributed it to a narrowing of the coronary vessels, and wrote the first clear description of atherosclerosis in medical history. He also described the pathology of cirrhosis of the man’s liver, which he found to be “desiccated and like congealed bran both in colour and substance”.

The dissection of the old man marked the beginning of five years of intense anatomical investigation, and in 1510–11

Leonardo seems to have collaborated with Marcantonio della Torre, the professor of anatomy at the University of Pavia.

Marcantonio provided ready access to human material, and Leonardo may have dissected up to 20 corpses that winter. He concentrated on the bones and muscles, analysing their structure in purely mechanical terms, and the results were spectacular (Fig. 2). Perhaps encouraged by the professional anatomist, Leonardo illustrated every bone except those of the skull, and most of the major muscle groups. The completion of his treatise was within reach, and on one drawing he wrote: “This winter of 1510 I believe I shall finish all this anatomy”.

But it was not so. In 1511, Milan descended into military turmoil. Marcantonio died of

the plague, and Leonardo retired to the country villa of his assistant, Francesco Melzi.

With the loss of his supply of human material, Leonardo reverted to the study of animal anatomy — most impressively the ox’s heart, which differs little in structure from that of a human. Leonardo described the ventricles and atria with great accuracy, and analysed the structure and functioning of the valves in minute detail. In a brilliant experiment, he made a glass model of the aortic valve (Fig. 3), through which he pumped water mixed with grass seeds to study the vortices in the widening of the aortic root. He deduced that these vortices were crucial in closing of the valve — a finding that was confirmed only in the twentieth century.

Leonardo had an almost perfect understanding of the physiology of the human heart. But he had no inkling of the circulation of the blood, and the existence of one-way valves was incompatible with the ancient belief that the heart simply churned blood in and out of the ventricles, thus generating heat and ‘vital spirit’. Unable to reconcile what he had observed with what he believed to be true, Leonardo reached an impasse. He became trapped in describing the motion of the blood through the valves in ever more detail. And there, it seems, his anatomical work came to an end.

There is no sign that Leonardo attempted to collate his research for publication. On his death in 1519 he left his papers to Melzi, and although the anatomical studies were mentioned by all Leonardo’s early biographers, their dense and disorganized content was barely comprehended. Unpublished, the studies were effectively lost to the world. Elsewhere, anatomical exploration gained pace, and in 1543 Andreas Vesalius published his epochal *De humani corporis fabrica* (*On the fabric of the human body*), a model of what Leonardo’s treatise could have been.

The 150 surviving sheets of Leonardo’s anatomical studies reached England in the seventeenth century and eventually made it into the Royal Collection, bound into an album with 450 of his more artistic drawings. But it was not until 1900 that they were finally published and understood. By then, their power to affect the progress of anatomical knowledge had long passed.

Leonardo’s paintings changed the course of European art; but his anatomical investigations, the finest of their age, were essentially unknown. This exhibition will give an unprecedented chance to assess this most paradoxical of scientists. ■

Martin Clayton is Senior Curator of Prints and Drawings at The Royal Library, The Royal Collection, Windsor Castle, Windsor, Berkshire, SL4 1NJ.
e-mail: martin.clayton@royalcollection.org.uk



Figure 3 | The findings recorded in *The aortic valve* (c.1512–13) went unconfirmed for more than 400 years.

EVOLUTION

Work–life balance

John Whitfield gauges a study that tackles the evolutionary conflicts behind workplace inequity.

My grandmother was a cleaner; my mother was a medical doctor. That is a single example of how, in the course of a few decades, women's educational and professional opportunities have increased vastly. As of 2010, about 60% of US women were in the workforce, making up nearly half of all workers; the picture for most other developed nations is similar. Women can do just about any job, and in many countries overt discrimination is illegal.

But other statistics tell a different story. In 2010, women made up just 1 in 7 company board members and 1 in 40 chief executives of Fortune 500 companies. And women still earn less than men for the same jobs. In the United States and Britain, a female lawyer earns, on average, about three-quarters of her male counterpart's salary. For newly qualified US medical doctors, the pay gap is widening: 17% in 2008, up from 12.5% a decade before.

In *The War of the Sexes*, Paul Seabright, an economist working at his subject's boundary with evolutionary biology, argues that human evolution can help to explain the inequities of today's white-collar, Western workplace. It is a story well worth hearing, but it is incomplete — partly because there is little decisive evidence for what causes gender gaps in employment, and partly because of some curious authorial choices.

The first half is an excellent primer on why males and females have evolutionary cause to behave in different ways. Sperm, Seabright explains, are cheap and plentiful; eggs, the opposite. This tends to make females choosy about their mates, and males less so. So males must compete for female attention by fighting among themselves, or by seducing prospective mates with advertisements of quality, such as gifts of food. Such differences in sexual supply and demand create conflicts of interest and incentives for deceit.

Paradoxically, our cooperative natures exacerbate these conflicts. Collaboration in child-rearing is key in humans because of the long childhood needed to grow a big brain. Men can provide resources such as meat, but in return, fathers and husbands have sought to control what (or, crudely but more accurately, who) women did.

In the second half of the book, Seabright investigates how such conflicts might explain the dearth of women at the top of the career ladder. Using evidence drawn mostly from the upper echelons of business,

Seabright suggests two causes of inequality.

First, anyone who takes a career break suffers for it. Even years after returning to work, he or she can expect to be paid less than a colleague who stayed put. So, Seabright argues in a claim that is likely to be controversial, there is no systematic discrimination



Despite a rise in professional opportunities for women, many are still paid less than men.

against women here — it is just that they tend to sacrifice office-time for child care. The cost of this, he says, results from a “signalling trap”: to get to the top, an employee must work all hours, not because it gets the job done, but because bosses take such behaviour as a sign of quality and commitment. Like a peacock investing in a huge tail to show off his good genes at the cost of his flying ability, this is hugely inefficient; but anyone who opts out unilaterally pays a disproportionate cost.

Second, Seabright argues that women are disadvantaged by networking behaviour. Each sex, for example, prefers to network with its own kind. In a male-led workplace, therefore, men's social and professional networks tend to overlap more than women's, so women are less likely to find professional opportunities. The evidence for this is more preliminary than that for the effect of career

breaks, but it points in the right direction. If you look in executive directors' pay packets, for example, women seem to benefit less than men from being well connected.

Throughout the book, Seabright is terrific company — entertaining and convincing. In a debate in which both sides tend to fit the evidence to their ideology, he lets the data take the lead. His recommendations for a more equal workplace, such as compulsory sex balance on job shortlists and mandatory paternity leave, are sensible and modest. Yet as a whole, *The War of the Sexes* doesn't quite hang together.

The first half shows that natural selection has given men and women different priorities, and so different strategies. But I struggled to make the leap from the economics of sex to the economics of work, including the modern pay gap. I would have liked to see a chapter stopping off midway between prehistory and the present, to look at how developments such as agriculture affected men and women. As Seabright notes, the biological similarities between the sexes show that women's “subordinate and dependent” state must be a relatively recent development — in which case, it seems odd to rely so heavily on the distant past.

In particular, the missing 10,000 years leave Seabright nowhere to tackle sexism and bullying. He assumes that people are rational and well-meaning. There is little discussion of cultural barriers to equality, such as male bosses' desire to keep women subordinate and dependent even if it costs them money. This is where those at the sharp end of the problem lay the blame: last month, a survey of women in UK banking found that three-quarters think that the biggest hurdle to equality is the attitude of senior male managers.

If this were a book by a science writer, I would no doubt be moaning that it sought to explain all of human life by shoe-horning it into one grand idea. Because it is a book by an academic, I'm moaning that it undercoheres, that the circumspection and caveats impede its argument. This is a dispatch from a field in its infancy, and the gaps in its thesis reflect the holes in our knowledge. But with people such as Seabright working to fill them, we can at least be optimistic that we will eventually understand what causes inequality, if perhaps less positive that we can fix it. ■

John Whitfield is a writer based in London. His latest book is *People Will Talk*. e-mail: j.a.whitfield@gmail.com

The War of the Sexes: How Conflict and Cooperation have Shaped Men and Women from Prehistory to the Present

PAUL SEABRIGHT
Princeton University Press: 2012. 256 pp.
\$24.95, £16.95

Correspondence

Keep a way open for tailored treatments

Two documents with crucial implications for the future of personalized medicine were released last month: a report from the US Institute of Medicine (IOM) emphasizing the importance of open data sharing (see go.nature.com/hofgoc) and a US Supreme Court decision ruling against two patents (see go.nature.com/lj9kyl), which could have the opposite effect.

This creates a quandary for companies aiming to pursue personalized treatments or methods, because those that adopt the IOM's advice to share their information now face greater hurdles to protecting their contributions through patents.

Companies should not be discouraged from exploring personalized medicine — to do so would deny society the benefit of tailored therapies. Likewise, firms that prefer to keep personalization algorithms secret will hinder scientific progress.

Instead, companies could sell value-added products that use open algorithms and statistical models — much like Google and IBM, for example, which have embraced open-source software.

To deliver effective treatments to patients, we need to find a way to reward companies for investing in personalized medicine while allowing the scientific community to advance knowledge rapidly.

Jeffrey T. Leek, Roger D. Peng
Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.
jtleek@gmail.com

R. Reeves Anderson *Arnold & Porter LLP, Washington DC, USA.*

Boost water safety in rural China

The world is well on its way to meeting the Millennium Development Goal to halve by

2015 the proportion of people who in 1990 had no sustainable access to safe drinking water (*Nature* **483**, 128–129; 2012). But more resources are still needed to improve water quality in China, particularly in rural areas.

According to a report released last month by the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), China contributed an estimated 457 million people to the 2 billion who by 2010 had access to improved water sources.

However, these 'improved' sources (a piped supply or borehole, for example, as opposed to a stream or unprotected well) do not always supply safe water, as judged by China's national standards for water quality.

Although 85% of China's rural population (605 million people) had access to improved water sources in 2010, only around one-quarter of these sources would be considered safe by national standards (R. Zhang *et al.* *J. Environ. Health* **26**, 3–5; 2009). By comparison with 1990 figures, this implies that only about 35 million people in rural China had gained access to safe water by 2010. For piped urban water, however, almost 91% of the water supply complied with chemical and microbial standards (J. Zhang *et al.* *J. Shenyang Agric. Univ.* **34**, 460–463; 2003), so roughly 295 million urban Chinese gained access to safe water over the same period.

The total Chinese population with access to safe water is therefore 330 million at most. This is about 127 million fewer people than estimated by the WHO/UNICEF report, a figure that is roughly equivalent to the population of Japan.

Hong Yang, Jim A. Wright
University of Southampton, UK.
hongyanghy@gmail.com
Stephen W. Gundry *University of Bristol, UK.*

Soil remedies for small-scale farming

In considering how best to improve poor agricultural soils in parts of Africa and improve crop productivity (*Nature* **483**, 525–527; 2012), micro-dosing would be a practical and inexpensive way to balance soil nutrients.

Developing countries tend routinely to use fertilizers containing only nitrogen, but a balanced supply of nutrients is essential for a crop to flourish. Urea application doubled in Africa during 1975–2005, causing an increase in the ratio of applied nitrogen to phosphorus and potassium (see go.nature.com/eanpxq). Although urea is cheap and its use is tempting to poor farmers in need of a rapid return on investment, it selectively depletes soil phosphorus, potassium and micronutrients. Ultimately, this imbalance will render the nitrogen ineffective.

Micro-dosing soil with small quantities of correctly proportioned 'NPK' fertilizer would significantly increase crop yields for smallholder farmers who are unable to afford large-scale applications. This simple technique involves dispensing measured amounts of the fertilizer using a bottle cap, for example, into planting holes during sowing. The returns will more than justify the small investment.

Marijn van der Velde, Linda See, Steffen Fritz *International Institute for Applied Systems Analysis, Laxenburg, Austria.*
velde@iiasa.ac.at

Getting a handle on biological data

We invite interested parties to help establish an infrastructure to improve the accessibility of the ever-increasing volumes of biological data. Our objective is to prepare a white paper this autumn for consideration by

the European Commission's Horizon 2020 funding call in 2013, among others.

New paradigms are needed to manage and process biological data. Good metadata descriptions for each data set will help users to locate the required facts and to make the best use of biological information.

Other infrastructural projects already exist, including the International Nucleotide Sequence Database Collaboration, the Catalogue of Life and the Global Biodiversity Information Facility, but they do not yet enable access to the broad swathes of data currently available.

Members of the community who are involved in generating and using biological data need to establish priorities so that funders can then make targeted grant calls, similar to the ten-year programme reviews established in the physical sciences.

We have initiated a community consultation as an open document (see go.nature.com/tma6gm) that describes a series of topics. Volunteers are invited to open similar documents for each topic, or to create a new topic, by adding their name and the document URL to the main index. A public meeting on the initiative will be held in Brussels on 17 July this year.

Dave Roberts *The Natural History Museum, London, UK.*
dmr@nomencurator.org
Alex Hardisty *Cardiff University, Cardiff, UK.*

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

Search for the compass needles

The idea that bird orientation is guided by magnetic-sensing structures in the animals' beaks has been challenged by the suggestion that the iron-containing cells are macrophages, which have no link to the brain. [SEE LETTER P.367](#)

HENRIK MOURITSEN

Most humans are hopeless navigators unless they have a map or the Global Positioning System to hand. By contrast, it is well known that many animals, including birds^{1,2}, use information from Earth's magnetic field to orient themselves, even over thousands of kilometres. But how do they sense the field? Iron-mineral-based structures within nerve cells in birds' upper beaks were previously proposed³ to act as one of two main magnetic sensors in birds, the other being a light-based system in the retina^{4,5}. However, on page 367 of this issue, Treiber *et al.*⁶ report that these iron-mineral deposits are in fact located inside macrophages, a kind of immune cell that is also involved in iron homeostasis⁷.

When people are asked how they think birds can sense magnetic fields, most suggest that birds must have small 'compass needles' somewhere in their body. The iron-containing minerals magnetite (Fe_3O_4) and maghemite (Fe_2O_3), which are found in various parts of a bird's body, are particularly well suited to sensing magnetic fields. However, these minerals can be relevant to magnetic sensing only if they are associated with nerve tissue and found in the same location in all individuals of a given species. In a landmark paper in 2003, Fleissner *et al.*³ reported that structures containing magnetite 'spherules' and maghemite 'platelets' are consistently present at six specific locations along the upper beaks of pigeons (Fig. 1a). These researchers suggested^{3,8} that the structures are magnetic sensors located in dendrites — nerve-cell endings. The structures included a large substructure that the researchers described^{3,8} as a vesicle.

If these structures sense the magnetic field, this information must be transmitted to the brain via the ophthalmic branch of the trigeminal nerve, as this is the only nerve that enters a bird's upper beak. Indeed, when this nerve is cut, pigeons trained to detect a strong magnetic field can no longer perform this task⁹. Furthermore, in European robins, which are migratory birds, neurons in the brain regions that receive direct input from the trigeminal nerve are activated by a changing magnetic-field stimulus, and this activation disappears if the trigeminal nerve is cut or if the magnetic

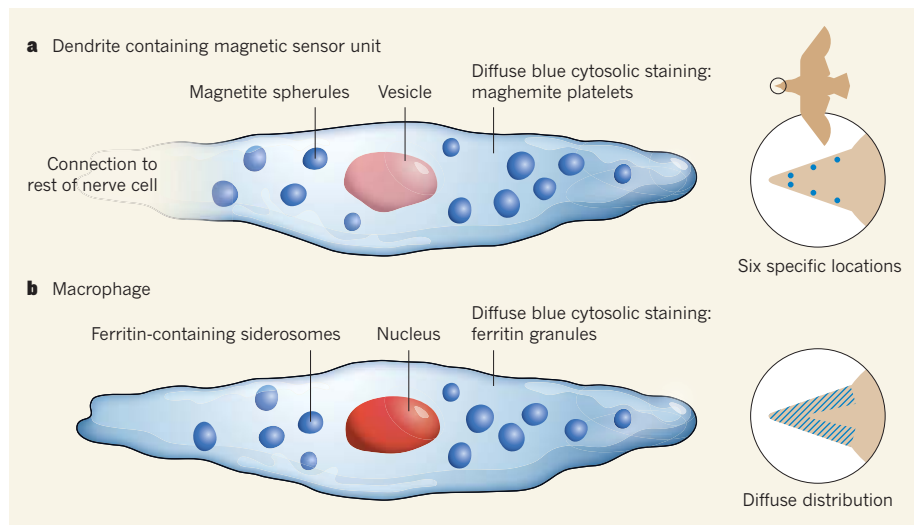


Figure 1 | Ironing out magnetoreception. **a**, Fleissner and colleagues^{3,8,11} previously proposed that birds' beaks contain dendrites (nerve-cell endings) harbouring iron-rich structures that can sense Earth's magnetic field and transmit this information to the brain, thereby helping the bird to navigate. By using electron and light microscopy to observe tissue sections stained with the iron-marking dye Prussian blue (blue colour), the authors identified magnetite-containing 'spherules', maghemite-containing 'platelets' (too small to depict) and a large substructure that they defined as a vesicle. They classified the structures as being clustered at six defined points along the upper beak. **b**, Treiber *et al.*⁶ now propose that these iron-mineral deposits are instead located in macrophages, immune cells that have a type of organelle (siderosomes) containing the iron-storage protein ferritin, and that the substructure is the cell's nucleus. The authors have found that these iron-containing cells are diffusely distributed throughout the beak and in other parts of the bird's body.

field is removed¹⁰. Therefore, even though there was no direct proof for their function, the iron-mineral structures described by Fleissner and colleagues^{3,8,11} have been generally assumed to be the primary magnetic receptors associated with the trigeminal magnetic sense in birds.

Treiber *et al.*⁶ challenge the interpretations of Fleissner and colleagues^{3,8,11}. On the basis of data that they collected from almost 200 pigeons, Treiber *et al.* propose that the iron-mineral-containing structures reported by Fleissner *et al.*³ are macrophages (Fig. 1b), not magnetic receptor cells. They show that the structures are located in variable quantities at inconsistent locations along the upper beak, and also elsewhere in the pigeon's body, such as in the respiratory tract and skin, and that they are not associated with nerve tissue⁶. They also suggest that the spherules, which are strongly stained by the iron-marking dye Prussian blue, are in fact organelles called siderosomes, which

contain the iron-storing protein ferritin, and that the large, round substructure is a nucleus, not a vesicle⁶.

Whose interpretation is correct? The answer to this question will be extremely important for magnetic-sense research. To form my own opinion, I examined many of the original microscope slides that make up the basis of Treiber and colleagues' report⁶, and compared them with my previous observations of some of the original slides that contributed to Fleissner and colleagues' studies^{3,11}. In my view, a significant proportion of the iron-mineral-containing structures found by Treiber *et al.*⁶ seem to be identical to those previously reported^{3,8,11}, even though some of the structures classified as such by Treiber *et al.* would, I believe, be regarded as artefacts by Fleissner and colleagues. However, even when I consider only those structures that look identical, these certainly seem to occur at many more

than six specific locations along the beak. Indeed, Treiber and colleagues' laborious quantification shows that the occurrence of the iron-mineral-containing structures is highly variable between individual birds⁶. By contrast, Fleissner and colleagues' claim^{3,8} of six specific locations is not supported by quantitative data. This lack of consistent distribution makes it highly unlikely that the iron-rich cells are part of a magnetic sensory system.

The most defining difference between macrophages and dendrites is that macrophages have a nucleus and dendrites do not. Treiber and colleagues' data and original slides convincingly show that the iron-containing structures are nucleated and that there is no regular co-localization between Prussian-blue staining and nerve fibres. The structures, therefore, cannot be located in dendrites. Furthermore, the authors report⁶ 95–99% co-localization of staining with Prussian blue and a macrophage marker, strongly indicating that most, but maybe not all, of the described structures are macrophages.

In conclusion, I find that serious doubt has been raised about the original proposals^{3,8,11}

and that the burden of proof now lies with Fleissner and colleagues, if they still think that their structures are potential magnetic sensors. However, it is important to stress that Treiber and colleagues' results cannot exclude the possibility that there are iron-mineral-based sensors somewhere in the upper-beak region of pigeons. Only relatively few sensors may be needed. In fact, a magnetite-based sensory cell might contain only a few magnetite crystals¹² and thus evade detection by Prussian-blue staining¹², the method used by both groups.

The implications of this new report are dramatic. The evidence supporting a magnetic compass embedded in birds' visual systems^{4,5} remains unaffected, but we are left with only two studies^{9,10} that clearly implicate the ophthalmic branch of the trigeminal nerve, although not necessarily iron-related structures, in magnetic sensing in birds. Therefore, for now, we have evidence supporting only a trigeminal-nerve-related magnetic sense in birds, not an iron-mineral-related magnetic sense. Both the biological function and basic sensory origin of the trigeminal magnetic sense are now unknown. Understanding the

magnetic senses in animals indeed remains a formidable scientific problem. ■

Henrik Mouritsen is at the Research Center for Neurosensory Sciences and the Institut für Biologie und Umweltwissenschaften, Carl-von-Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany.
e-mail: henrik.mouritsen@uni-oldenburg.de

1. Wiltchko, W. & Wiltchko, R. *Science* **176**, 62–64 (1972).
2. Cochran, W. W., Mouritsen, H. & Wikelski, M. *Science* **304**, 405–408 (2004).
3. Fleissner, G. *et al. J. Comp. Neurol.* **458**, 350–360 (2003).
4. Ritz, T., Thalau, P., Phillips, J. B., Wiltchko, R. & Wiltchko, W. *Nature* **429**, 177–180 (2004).
5. Zapka, M. *et al. Nature* **461**, 1274–1277 (2009).
6. Treiber, C. D. *et al. Nature* **484**, 367–370 (2012).
7. Wang, J. & Pantopoulos, K. *Biochem. J.* **434**, 365–381 (2011).
8. Fleissner, G., Stahl, B., Thalau, P., Falkenberg, G. & Fleissner, G. *Naturwissenschaften* **94**, 631–642 (2007).
9. Mora, C. V., Davison, M., Wild, J. M. & Walker, M. M. *Nature* **432**, 508–511 (2004).
10. Heyers, D., Zapka, M., Hoffmeister, M., Wild, J. M. & Mouritsen, H. *Proc. Natl Acad. Sci. USA* **107**, 9394–9399 (2010).
11. Falkenberg, G. *et al. PLoS ONE* **5**, e9231 (2010).
12. Walker, M. M. *et al. Nature* **390**, 371–376 (1997).

ELECTRONICS

Carbon nanotubes finally deliver

A carbon-nanotube transistor has been made that performs better than the best conventional silicon analogues. The result propels these devices to the forefront of future microchip technologies.

FRANZ KREUPL

Every time you turn on your computer to browse the Internet or write a presentation, billions of tiny switches called transistors orchestrate a dance of electrons in the computer's processor that is tuned to the rhythm of the underlying software. By letting electric current flow or not, the transistors, which are made of silicon, create the states (the 'zeros' and 'ones') on which our Boolean computer world is currently based. Writing in *Nano Letters*, Franklin *et al.*¹ describe how they have created a transistor based on carbon that promises to outperform its silicon counterparts in future electronic circuitry.

The improvement in the look and feel of computers over the past decades — as they turned from clunky, nerdy objects into stylish high-tech gadgets — is partly due to the fact that silicon transistors have shrunk in size from tens of micrometres to about 20–30 nanometres. However, further scaling down of this technology is becoming more and

more challenging. That is why researchers have been looking for a successor to silicon transistors that would allow further miniaturization — hitherto with limited success.

Franklin and colleagues' study represents progress on this front. The authors show that single-walled carbon nanotubes (SWCNTs) — semiconductors that have been under close scrutiny for more than a decade — can operate as excellent switches at molecular dimensions of only 9 nm (Fig. 1). This is less than half the size of the leading silicon technology. And the authors' transistor delivers more current and requires less operating power than do silicon-based competitors of equivalent size. This is the first experiment to clearly demonstrate that silicon could be replaced with carbon material in future semiconductor technology.

To operate in the same way as the usual, larger silicon transistors, SWCNT transistors must show what is known as 'current saturation' at low applied voltage. This means that the devices' output current should hardly be affected if the applied voltage is varied. Exactly this behaviour

is demonstrated by the authors' device¹ at a supply voltage of less than 0.4 volts, lower than the 0.64 V that the roadmap² of the semiconductor industry postulates for a device size of 9 nm, scheduled to reach the market in 2022.

Early attempts³ to look at the scalability of SWCNT transistors revealed that an 18-nm-long SWCNT device showed no current saturation — the current output was simply proportional to the applied voltage. But it was unclear whether this was an inherent feature of ballistic electron transport (in which no electrons are scattered) or due to the electrostatic design of the device⁴. Franklin *et al.*^{1,5} investigated this issue in more detail, and improved the current-saturation effect by employing a better electrostatic design than that used previously.

In an earlier paper⁵, the authors were able to demonstrate that the use of a 10-nm-thick insulating layer of hafnium oxide, in place of the layer of silicon oxide used in the 18-nm SWCNT device, gives a much stronger electrostatic coupling between the device's gate electrode and the SWCNT, and enables a 20-nm-long SWCNT transistor to be created that has good current-saturation behaviour. However, a 15-nm-long SWCNT device, which was fabricated in the same batch of transistors, did not show this promising signal⁵ because the electrostatic coupling was apparently weakened by the shorter distance between the device's source and drain electrodes.

The transistor that Franklin *et al.* now report is based on an electrostatic design involving a 9-nm-long channel made of a SWCNT and a hafnium oxide insulating layer that is only 3 nm thick. The reduced thickness of the insulating layer produced much stronger

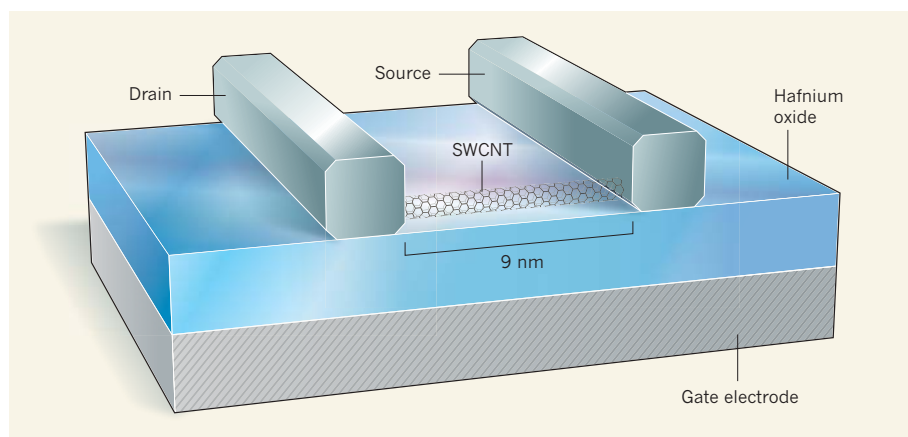


Figure 1 | A carbon-nanotube transistor. Franklin and colleagues' transistor¹ consists of a 9-nanometre-long channel made of a single-walled carbon nanotube (SWCNT), which is a single layer of carbon atoms arranged on a honeycomb pattern and rolled into a seamless cylinder about 1 nm across. When a voltage is applied to the transistor's gate electrode, electric current flows from the source electrode to the drain electrode. A 3-nm-thick layer of hafnium oxide allows strong electrostatic coupling between the gate and the SWCNT.

electrostatic coupling between the short SWCNT channel and the gate electrode than is achieved with a thicker oxide. What's more, the authors show that the strong electrostatic coupling enables the device to fulfil another requirement of an electronic switch: it can 'turn off' the current to values that are low enough for applications.

Although transistors operate as switches, they can never turn the current off completely. The question is, how much can the current be reduced? In Franklin and colleagues' transistor, the off-current was 10,000 times smaller than

when the channel was open. And it may be possible to suppress the current even further by deploying a better electrostatic design than that used by the authors. Although carbon nanotubes were predicted a decade ago to be the way forward for microelectronics, direct experimental proof that they can be used to make transistors at molecular scales was missing until now.

Franklin and colleagues' transistor is good news, because power consumption has an ever-increasing role in electronics, and the device's ability to operate at small supply voltages should mitigate the demand in energy

consumption in future microchips. However, the road to a real application is bumpy and beset with challenges. Some of these challenges, such as the reliability and variability of the transistors, could be solved by sophisticated engineering methods. But others, including the creation of pure batches of semiconducting nanotubes, are of a more fundamental nature. SWCNTs can exist in 'flavours', each of which displays different electronic properties. If billions of SWCNTs are needed on a chip, they must all have the same flavour, which is currently difficult to attain.

Nonetheless, if the progress^{6,7} made in the field of nanotubes during the past decade is anything to go by, the prospects for solving these fundamental problems look good. Franklin and colleagues' results will certainly stimulate further research to unlock more secrets of carbon nanotubes. ■

Franz Kreupl is in the Department of Hybrid Electronic Systems, Technische Universität München, 80333 München, Germany.
e-mail: franz.kreupl@tum.de

1. Franklin, A. D. *et al.* *Nano Lett.* **12**, 758–762 (2012).
2. *International Technology Roadmap for Semiconductors 2011 Edition: Front End Processes* available at go.nature.com/qykuh1 (2011).
3. Seidel, R. V. *et al.* *Nano Lett.* **5**, 147–150 (2005).
4. Ferry, D. K., Gilbert, M. J. & Akis, R. *IEEE Trans. Electron Dev.* **55**, 2820–2826 (2008).
5. Franklin, A. D. & Chen, Z. *Nature Nanotechnol.* **5**, 858–862 (2010).
6. Hata, K. *et al.* *Science* **306**, 1362–1364 (2004).
7. Kang, S. J. *et al.* *Nature Nanotechnol.* **2**, 230–236 (2007).

rapidly quenched³ and alga-like microfossils⁵. Indirect evidence, meanwhile, extends the record of terrestrial oceans to as early as 4.4 billion years ago⁶. So how can this be explained?

Solutions to the paradox fall into two general categories: those contending that Earth's atmosphere retained heat more efficiently in the past than it does now, for example because of increased concentrations of greenhouse gases, and those arguing that the albedo, or reflectance, of Earth was lower in the past, perhaps because there were fewer clouds and/or less ice. Most models used to explain the paradox are purely theoretical, and were designed to highlight the conditions necessary to solve it. Unfortunately, few observational constraints to support or refute these models have been identified, and those that have been proposed^{7–10} tend to be controversial.

Som *et al.*³ have implemented a clever approach — first attempted¹¹ in 1851 by a pioneer of geology, Charles Lyell — to determine the density of the atmosphere early in Earth's history. This information is crucial for assessing whether greater concentrations of greenhouse gases (such as carbon dioxide¹²), or of other gases (such as nitrogen¹³) that amplify the effects of greenhouse gases, could explain the faint young Sun paradox.

GEOSCIENCE

Fossil raindrops and ancient air

An analysis of fossil imprints of ancient raindrops suggests that the density of the atmosphere 2.7 billion years ago was much the same as that today. This result casts fresh light on a long-standing palaeoclimate paradox. SEE LETTER P.359

WILLIAM S. CASSATA & PAUL R. RENNE

When a main-sequence star such as our Sun ages, its inner core becomes denser and the temperature at which its hydrogen is fused to helium increases. As a result, the Sun is currently more luminous, and delivers more energy to Earth's surface, than in the past — its energy output around 2 billion years ago is inferred¹ to have been less than 85% of that today. Such a faint Sun should not have been able to warm Earth's surface above the freezing point of water², yet the geological record indicates that liquid water abounded at that time. The apparent contradiction between

theory (sub-freezing Earth surface temperatures) and observation (liquid water) is known as the 'faint young Sun' paradox. On page 359 of this issue, Som *et al.*³ address this paradox using seemingly unlikely evidence: fossilized imprints left by ancient raindrops*.

There are ample indications of liquid water under the faint Sun. In South Africa, for example, exposed rocks more than 3 billion years old contain features associated with flowing or standing water, including sedimentary deposits that preserve ripple marks and mud cracks⁴, glassy 'pillow' lavas that were

*This article and the paper under discussion³ were published online on 28 March 2012.

Specifically, the authors inferred the velocity of falling ancient raindrops from the geometry of fossilized raindrop-impact marks preserved in a 2.7-billion-year-old sedimentary rock from South Africa (Fig. 1). The atmosphere exerts a drag on raindrops such that they typically fall at a terminal velocity that is inversely proportional to the density of the atmosphere. Any difference between the inferred velocity of ancient raindrops and that of those that fall today may therefore reflect a change in atmospheric density. Of course, the imprint generated by a raindrop falling at a given velocity depends both on the size of the drop and on the nature of the substrate onto which it falls. To constrain these variables, the authors observed the size distributions of naturally occurring raindrops, and coupled this information with data from experiments in which they let water droplets fall onto volcanic ash — mimicking the conditions in which the fossil raindrops formed.

Som *et al.* conclude that the atmospheric density 2.7 billion years ago was probably 50 to 105% of that today. This finding immediately calls into question solutions to the faint young Sun paradox that invoke elevated concentrations of greenhouse gases, unless small increases of greenhouse-gas concentration were able to exert a large warming effect. It is also unlikely that higher concentrations of greenhouse-enhancing nitrogen could have caused the paradox, because concentrations of twice or more the present atmospheric abundance would have been required to provide sufficient warming in the presence of a modest increase in carbon dioxide¹³. Under such conditions, the atmospheric density would have been

greater than that predicted by the authors. It therefore seems that elevated concentrations of highly effective greenhouse gases, such as methane¹⁴, ethane¹⁵ and/or carbonyl sulphide¹⁶, may be required to explain the paradox, possibly in combination with moderately higher concentrations of less-effective greenhouse gases such as carbon dioxide¹⁴. A lower planetary albedo, caused by the reduction or absence of continental ice sheets, could also have contributed to warming.

Although raindrop size distributions associated with typical storms are well known, it is possible — albeit unlikely — that the ancient raindrops responsible for the fossilized imprints were unusually large. Small errors in the inferred size of the raindrops would result in significant errors in the atmospheric pressure predicted by Som and colleagues' method³. The accuracy of the method is further limited by lack of information about factors (such as moisture content) that would have affected the cohesiveness of the ash in which the fossil imprints were made; the cohesiveness can affect the morphology of impact craters. The atmospheric density 2.7 billion years ago could therefore have been more than twice that of the modern atmosphere if the circumstances under which the fossil imprints formed were unusual.

It is to be hoped that Som and colleagues' work will stimulate further studies of fossil raindrop imprints, including perhaps those originally observed by Lyell. In particular, it will be interesting to see whether coherent temporal trends in atmospheric pressure can be inferred from imprints in deposits of varying ages. With increasing recognition and analysis of such features in the geological record, it may be possible to establish a chronological record of atmospheric pressure on Earth throughout the past 3.5 billion years. Such a record would shed light on otherwise poorly constrained aspects of climate change deep in Earth's history. ■

William S. Cassata and Paul R. Renne are in the Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, California 94720, USA, and at the Berkeley Geochronology Center.
e-mails: cassata@berkeley.edu; prenne@bgc.org



Figure 1 | Solid evidence. Som *et al.*³ have analysed the fossilized imprints of raindrops, such as those shown here, to determine the atmospheric density 2.7 billion years ago. Rule, 5 cm.

1. Ribas, I. in *Solar and Stellar Variability: Impact on Earth and Planets* (eds Kosovichev, A. G., Andrei, A. H. & Rozelot, J.-P.) Ch. 1 (Cambridge Univ. Press, 2010).
2. Sagan, C. & Mullen, G. *Science* **177**, 52–56 (1972).
3. Som, S. M., Catling, D. C., Harnmeijer, J. P., Polivka, P. M. & Buick, R. *Nature* **484**, 359–362 (2012).
4. Ramsay, J. G. *Trans. Geol. Soc. S. Afr.* **66**, 353–401 (1963).
5. Schopf, J. W. & Barghoorn, E. S. *Science* **156**, 508–512 (1967).
6. Harrison, T. M. *Annu. Rev. Earth Planet. Sci.* **37**, 479–505 (2009).
7. Reinhard, C. T. & Planavsky, N. J. *Nature* **474**, <http://dx.doi.org/10.1038/nature09959> (2011).
8. Dauphas, N. & Kasting, J. F. *Nature* **474**, <http://dx.doi.org/10.1038/nature09960> (2011).
9. Goldblatt, C. & Zahnle, K. J. *Nature* **474**, <http://dx.doi.org/10.1038/nature09961> (2011).
10. Rosing, M. T., Bird, D. K., Sleep, N. H. &



50 Years Ago

Charles Darwin's Orchid Bank at Downe in Kent has recently been acquired as a nature reserve by the Kent Naturalists' Trust. It is a site of great scientific interest which shows a wide range of vegetation types ... and, as it is only sixteen miles from the centre of London, the need for conservation is urgent and considerable. It is also of unique historical importance, for Darwin is known to have carried out field studies there from his nearby home, Down House, where he lived for forty years ... Eleven of the thirteen species of orchid mentioned by Darwin as occurring within a mile of his house can still be found in the reserve.

From *Nature* 21 April 1962

100 Years Ago

The Mind of Primitive Man. By Franz Boas — There is a popular fallacy that racial antipathy is based on physiological foundations. But in so far as such antipathy is real, there is nothing physiological in its causation ... The author's discussion and explanation of the causes and results of variation within a race ... supply the most convincing theory that has yet appeared ... The ordinary view of the mental deficiencies of the "inferior races" is remorselessly criticised. The lowest savage *does* possess self-control. He is not improvident, but rather optimistic. He *can* concentrate his mind. He possesses originality. Savages who do not count beyond three or ten easily adapt their language and intellect to civilised methods of reckoning ... The point is that these civilised methods are not needed in the primitive state, where each man on a war-expedition is known by name, though the number of the troop may not be reckoned. Both in mind and in body there is little to choose between the ordinary barbarian and the civilised man.

From *Nature* 18 April 1912

- Bjerrum, C. J. *Nature* **474**, <http://dx.doi.org/10.1038/nature09962> (2011).
 11. Lyell, C. *Edinb. N. Phil. J.* **51**, 70–74 (1851).
 12. Kasting, J. F. *Precamb. Res.* **34**, 205–229 (1987).
 13. Goldblatt, C. *et al. Nature Geosci.* **2**, 891–896 (2009).

14. Sheldon, N. D. *Precamb. Res.* **147**, 148–155 (2006).
 15. Haqq-Misra, J. D., Domagal-Goldman, S. D., Kasting, P. J. & Kasting, J. F. *Astrobiology* **8**, 1127–1137 (2008).
 16. Ueno, Y. *et al. Proc. Natl Acad. Sci. USA* **106**, 14784–14789 (2009).

QUANTUM PHYSICS

Tunnelling across a nanowire

The observation of a phenomenon known as coherent quantum phase slip, across a nanowire in a superconducting system, paves the way for applications in quantum computing and metrology. [SEE LETTER P.355](#)

ALEXEY BEZRYADIN

Quantum mechanics is the most accurate theory of modern physics. It was originally formulated¹ in 1925 to describe microscopic particles such as electrons and atoms, but whether the theory is applicable to the macroscopic world of everyday objects has remained unclear. On page 355 of this issue, Astafiev *et al.*² demonstrate that quantum theory can describe the tunnelling of magnetic flux across a narrow segment of a superconducting loop. Taken as a whole, this device represents a macroscopic and complex system, involving — at the very least — probably many thousands of electrons.

Astafiev and colleagues² have created a type of superconducting quantum bit (qubit) that was first proposed by Mooij and Harmans³. The device's main element is a segment of a homogeneous nanowire within a closed superconducting loop (Fig. 1). The system operates by allowing quantum tunnelling of magnetic flux, into and out of the loop, across the nanowire. Such tunnelling preserves a form of quantum memory known as phase coherence and is called coherent quantum phase slip (CQPS). The device obeys the physics of macroscopic quantum systems and has implications for fundamental metrology and information technology.

The principle of quantum tunnelling⁴ posits that any microscopic particle has some chance of penetrating any wall, no matter how high the wall's associated energy barrier. But does this effect apply to macroscopic objects? On the basis of one modern interpretation of quantum mechanics⁵, quantum effects are applicable to any large system, even the Universe. But when applied to macroscopic objects, such a global theory would lead to paradoxical predictions — for example, Schrödinger's cat, which, according to the principle of quantum superposition, can be alive and dead at the same time.

However, another interpretation exists in

which large Schrödinger's cats do not occur because of a phenomenon known as spontaneous wavefunction collapse^{6,7}. Such a collapse destroys phase coherence and occurs spontaneously at a low rate. Therefore, experimental physicists are avidly testing the applicability of quantum theory to the macroscopic world⁸. To do this, they often start with tiny electronic devices, which display coherent quantum tunnelling between macroscopically distinct states^{9,10}. Such small-scale but macroscopic quantum systems may be called artificial atoms, given the discrete nature of their energy states, and can be used as qubits, the building blocks of quantum computation.

The Mooij–Harmans qubit, which Astafiev *et al.* now demonstrate², can exist either in a state in which the superconducting current in the loop flows clockwise or in a state that has an anticlockwise current. It can also be in a symmetric quantum superposition of these two states, which is described mathematically by the sum of the clockwise and anticlockwise states. What's more, the symmetric superposition state has a twin state of higher energy, called the antisymmetric state. This state is obtained by subtracting the anticlockwise state from the clockwise state. In their study, the authors prepared the qubit in the symmetric superposition state and, by shining microwave photons on the qubit, were able to make it switch between the two superposition states. To detect these states, they connected the qubit to the centre conductor of a microwave resonator system, and showed that the two states slow down electromagnetic waves differently.

In a Mooij–Harmans qubit, the transition from the clockwise state to the anticlockwise state is accompanied by tunnelling of magnetic flux, or quantum phase slip, across the nanowire. If the qubit's wire is continuous, the magnetic flux enters the loop by creating what is known as a phase-slip core. Such a core is similar to the normal core of a vortex of magnetic flux in a superconductor. But CQPS, in which the quantum history of the tunnelling

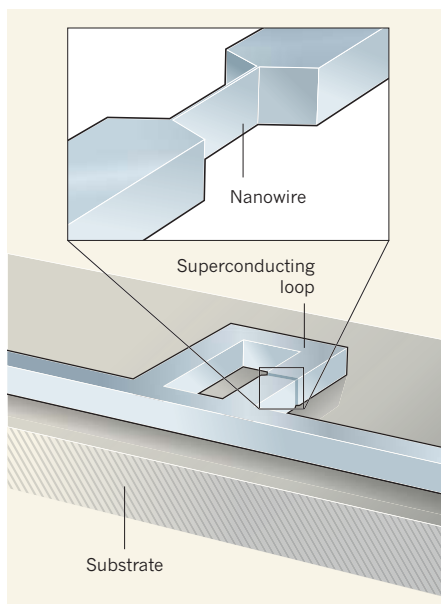


Figure 1 | The Mooij–Harmans qubit. The two-state quantum system, or qubit, demonstrated by Astafiev *et al.*² consists of a nanowire segment in a superconducting loop made of indium oxide and laid on a substrate. The strongly disordered electronic structure of indium oxide allows coherent quantum tunnelling of magnetic flux, known as coherent quantum phase slip, to occur across the nanowire. (Figure not drawn to scale.)

(the coherence) is preserved, is difficult to attain because the core dissipates energy, erases the quantum memory, causes decoherence and suppresses the tunnelling rate exponentially¹¹. As a result, only incoherent quantum phase slips had thus far been detected in homogeneous nanowires^{12,13}.

Previous observations of CQPS have involved the use of superconducting qubits made of nanowires interrupted by thin insulating junctions. Such junctions create tunnel barriers for the superconducting current, but help the magnetic flux to cross the wire and enter the loop because superconductors expel the magnetic field from their interior whereas insulators do not. Also, CQPS in these interrupted wires has no core. Therefore, the tunnelling rate is not strongly influenced by dissipation effects¹¹. Astafiev and colleagues' superconducting qubit² is the first of its kind in which CQPS is observed across an uninterrupted superconducting nanowire.

The observation of CQPS in a continuous nanowire was possible because the superconducting material from which the wire is made — indium oxide — has a strongly disordered electronic structure. Because of its oxygen atoms' tendency to bind or localize electrons, indium oxide is so disordered that it is not even described by the standard Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity. In fact, the material is almost an insulator. In BCS-type superconductors, all free electrons participate in a single, collective bound state

known as the BCS condensate. Such a state is like a huge superconducting 'molecule' of electrons, which moves without friction and so leads to superconductivity. In indium oxide, the BCS condensate is not one but many such molecules. The condensate is localized, meaning that it fractures into a system of BCS 'lakes' or 'droplets', which interact only very weakly with one another.

Astafiev and colleagues' work suggests that the presence of such lakes is crucial to the successful operation of the Mooij–Harmans qubit. In a superconducting wire that contains localized BCS-condensate lakes, the phase-slip core can pass between the lakes and thereby avoid any strong dissipation. Hence, instead of the conventional approach of making an insulating junction in the nanowire to observe CQPS, the researchers used a material (indium oxide) that is almost in an insulating state. This is a neat solution because, on the one hand, the wire has some phase-slip cores, but, on the other hand, dissipation effects are reduced.

The present study paves the way for testing theoretical predictions for nanowires supporting CQPS and for applications in fields such as quantum computing and quantum metrology. It has been proposed¹⁴ that CQPS could be used to produce a fundamental standard of current. One issue that remains to be investigated is the possibility — or perhaps impossibility — of attaining CQPS in nanowires that are less disordered than indium oxide, and, in

particular, in pure, single-crystal nanowires in which BCS theory is fully applicable and a well-developed phase-slip core exists¹⁵. Such systems should be compared with those that support coreless CQPS¹⁶, to quantify the decoherence introduced by the phase-slip core. ■

Alexey Bezryadin is in the Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801–3080, USA. e-mail: bezryadi@illinois.edu

1. Heisenberg, W. Z. *Phys.* **33**, 879–893 (1925).
2. Astafiev, O. V. *et al.* *Nature* **484**, 355–358 (2012).
3. Mooij, J. E. & Harmans, C. J. P. M. *New J. Phys.* **7**, 219 (2005).
4. Gamow, G. Z. *Phys.* **51**, 204–212 (1928).
5. Everett, H. III *Rev. Mod. Phys.* **29**, 454–462 (1957).
6. Bell, J. S. in *Schrödinger: Centenary Celebration of a Polymath* (ed. Kilmister, C. W.) 41–52 (Cambridge Univ. Press, 1987).
7. Tumulka, R. *Proc. R. Soc. Lond. A* **462**, 1897–1908 (2006).
8. Gerlich, S. *et al.* *Nature Commun.* **2**, 263 (2011).
9. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).
10. Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. *Nature* **406**, 43–46 (2000).
11. Caldeira, A. O. & Leggett, A. J. *Phys. Rev. Lett.* **46**, 211–214 (1981).
12. Bezryadin, A. J. *Phys. Condens. Matter* **20**, 043202 (2008).
13. Bezryadin, A. *Superconductivity in Nanowires — Fabrication and Quantum Transport* (Wiley, in the press).
14. Mooij, J. E. & Nazarov, Yu. V. *Nature Phys.* **2**, 169–172 (2006).
15. Golubev, D. S. & Zaikin, A. D. *Phys. Rev. B* **64**, 014504 (2001).
16. Manucharyan, V. E. *et al.* *Phys. Rev. B* **85**, 024521 (2012).

CIRCADIAN RHYTHMS

No lazing on sunny afternoons

In the laboratory, fruitflies rely on an internal clock to alternate activity with a midday nap and night-time sleep. Surprisingly, when outdoors, they follow temperature rather than the clock, and skip siestas. SEE LETTER P.371

FRANÇOIS ROUYER

Circadian clocks are biological mechanisms that allow living organisms to adapt their physiology and behaviour to day–night cycles. In particular, a neuronal circadian clock controls rhythms of rest and activity in many animal species, including the fruitfly *Drosophila melanogaster*. In laboratory conditions, this insect shows a bimodal activity profile, with maximal activity concentrated in the morning and evening. However, the artificial light–dark cycles used in lab experiments only roughly mimic the complex environmental changes associated with alternating day and night. Vanin *et al.*¹ report on page 371 of this issue that fruitflies indeed behave very

differently when exposed to more natural conditions. Moreover, the authors' results challenge the widely held idea that light and the known elements of the circadian clock are the main players that shape rest–activity rhythms.

Fruitflies show 24-hour sleep–wake rhythms that persist for weeks in laboratory conditions at constant room temperature, even in constant darkness. The insects' circadian clock relies on a feedback loop that involves four proteins — CLOCK, CYCLE, PERIOD and TIMELESS — that bind to specific regions of DNA and control the expression of many genes². Mutant fruitflies devoid of any one these four proteins lack circadian rhythms. Obviously, though, the constant conditions used in the lab are not suited to understanding the adaptive role of the

clock under natural day–night cycles.

Why do animals (and humans) need an internal clock in a cycling environment? Textbooks tell us that a clock allows organisms to anticipate environmental changes and therefore to prepare for such changes. And anticipating is exactly what fruitflies do in the laboratory (Fig. 1a). In light–dark conditions (12 hours of light and 12 hours of darkness, at constant room temperature), the insects show little activity at night and wake up a few hours before the lights are turned on. After their morning activity, they take an afternoon nap and then become progressively more active until the moment when the lights are switched off and sleep time starts. Such bimodal activity profiles are thought to restrict the fruitflies' behaviour in the wild to the times of the day (morning and evening) that have the most favourable temperature and humidity conditions. By contrast, clock-less mutant fruitflies do not anticipate laboratory light–dark cycles and show only 'startle' responses to light transitions, which demonstrates that a clock is required for these anticipatory activities.

To understand how the circadian clock affects the insects' behaviour in more natural conditions, Vanin *et al.*¹ placed the flies, together with activity-recording devices, outside the lab. The devices were shielded from direct sunlight but subject to natural variations of light and temperature. The authors did these experiments from spring to autumn in two geographical locations, Leicester (UK) and Treviso (Italy), and used clock-defective mutants and normal fruitflies.

These outdoor experiments brought big surprises. The researchers observed that normal fruitflies, as well as most mutants (including clock-less insects that lacked PERIOD or TIMELESS), displayed a prominent afternoon activity peak in addition to the morning and evening peaks (Fig. 1b). The afternoon activity was temperature-dependent: for each of the fruitfly strains that showed afternoon activity when the mean day temperature was 27 °C, only 50% of the insects did so at 20 °C. Some afternoon activity had been reported previously^{3–5} when temperature cycles were used in lab conditions. When the authors tested *per^S* mutant fruitflies, which express a defective PERIOD protein, leading to an earlier timing of evening activity, they found that the afternoon peak also took place earlier.

Vanin and colleagues' results¹ indicate that the afternoon activity is modulated by the circadian clock, and seem to exclude a simple heat-induced response. They also suggest that the insects are highly active in the middle of the day under natural conditions, a conclusion that

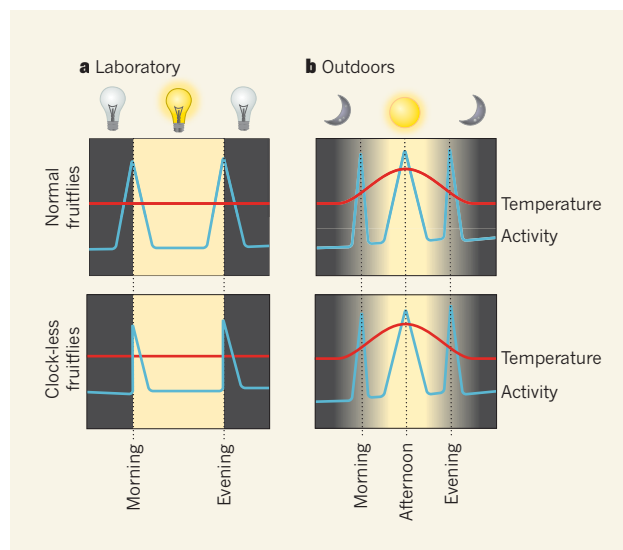


Figure 1 | Rest-activity rhythms in fruitflies. **a**, In laboratory-induced dark–light cycles (12 hours dark, 12 hours light and a constant temperature of 25 °C), normal fruitflies show morning and evening anticipatory activities, whereas clock-less mutant insects do not. **b**, When Vanin *et al.*¹ studied these processes outdoors (in natural light and temperature conditions), they found that normal fruitflies and clock-less mutants show similar morning and evening activities, as well as a prominent afternoon activity peak. The authors' results suggest that temperature is a key factor in shaping the flies' activity under outdoor conditions.

contrasts with results of previous lab experiments^{6,7} which showed that fruitflies prefer low to high light intensity and tend to show increased activity when simulated moonlight is applied at night. However, in the authors' outdoor experiments¹, the amount of activity restricted to the morning and evening twilights was always less than 25% of the total activity, and the insects did not tend to be more nocturnal around the time of a full moon. These results suggest that, in the wild, fruitflies are much more diurnal than crepuscular, in contrast to what lab experiments tell us.

Vanin *et al.* then carefully analysed the timing of morning and evening activities, which are driven apart when day length increases, and observed that the onset of morning activity was similar in normal and clock-defective insects. However, for all the fruitfly strains tested, except for clock-less mutants lacking TIMELESS, the higher the mean temperature, the earlier was the morning wake-up. In other words, temperature — and not the clock — seems to tell the insects when to become active. When the authors used a cover to prevent the fruitflies from seeing twilight, the temperature-dependent onset of the morning activity was delayed. Therefore, twilight shifts the temperature-dependent start of morning activity to an earlier time. Why TIMELESS-lacking insects do not similarly respond to temperature remains unknown, but it has been shown⁸ that these mutants show altered diapause, a seasonal overwintering response that also requires a coordinated reaction to light and temperature.

What about evening activity? The authors

observed that, on cold days (mean temperature below 20 °C), the flies' evening activity started in the middle of the day, between the time of highest light intensity and that of highest temperature about 2 hours later. On hot days, by contrast, they found that, the higher the temperature, the later was the onset of evening activity. Temperature, therefore, sets the phase of evening activity. Of note, evening onset was earlier in *per^S* mutants, which indicates that — in contrast to the onset of morning activity — the beginning of evening activity is controlled by the circadian clock.

The outdoor experiments performed by Vanin *et al.*¹ shake up our assumptions about the timing of rest-activity rhythms in the fruitfly. The presence or absence of a circadian clock seems to have less effect on the insects' activity than expected, although the clock does regulate afternoon and evening activities. As light and temperature can show abrupt and unexpected variations during the day, a key function of the circadian clock might be to act as a buffering system to keep day activity stable.

Ironically, although laboratory studies^{9,10} indicate that the same circadian-clock neurons can drive both morning activity and sustained circadian rhythms, the clock does not seem to influence morning activity. These neurons might therefore have roles in addition to triggering morning activity. It will be important to see whether other circadian-modulated processes in the fruitfly, such as the emergence of young adults from pupae (a life stage between larva and adult), obey the rules revealed by Vanin and colleagues' experiments. In any case, the present work is a starting point for bringing the natural environment into the lab and thus for beginning to tackle the cellular and molecular basis of real-life circadian behaviour. ■

François Rouyer is at the *Institut de Neurobiologie Alfred Fessard, CNRS UPR 3294, 91198 Gif-sur-Yvette, France.*
e-mail: francois.rouyer@inaf.cnrs-gif.fr

1. Vanin, S. *et al.* *Nature* **484**, 371–375 (2012).
2. Hardin, P. E. *Adv. Genet.* **74**, 141–173 (2011).
3. Tomioka, K., Sakamoto, M., Harui, Y., Matsumoto, N. & Matsumoto, A. *J. Insect Physiol.* **44**, 587–596 (1998).
4. Busza, A., Murad, A. & Emery, P. *J. Neurosci.* **27**, 10722–10733 (2007).
5. Glaser, F. T. & Stanewsky, R. *Curr. Biol.* **15**, 1352–1363 (2005).
6. Rieger, D. *et al.* *J. Biol. Rhythms* **22**, 387–399 (2007).
7. Bachleitner, W., Kempinger, L., Wülbeck, C., Rieger, D. & Helfrich-Förster, C. *Proc. Natl Acad. Sci. USA* **104**, 3538–3543 (2007).
8. Tauber, E. *et al.* *Science* **316**, 1895–1898 (2007).
9. Stoleru, D., Peng, Y., Agosto, J. & Rosbash, M. *Nature* **431**, 862–868 (2004).
10. Grima, B., Chélot, E., Xia, R. & Rouyer, F. *Nature* **431**, 869–873 (2004).

Deformation cycles of subduction earthquakes in a viscoelastic Earth

Kelin Wang^{1,2}, Yan Hu^{2†} & Jiangheng He¹

Subduction zones produce the largest earthquakes. Over the past two decades, space geodesy has revolutionized our view of crustal deformation between consecutive earthquakes. The short time span of modern measurements necessitates comparative studies of subduction zones that are at different stages of the deformation cycle. Piecing together geodetic ‘snapshots’ from different subduction zones leads to a unifying picture in which the deformation is controlled by both the short-term (years) and long-term (decades and centuries) viscous behaviour of the mantle. Traditional views based on elastic models, such as coseismic deformation being a mirror image of interseismic deformation, are being thoroughly revised.

A subduction fault may stay ‘locked’ for centuries and then suddenly slip to cause an earthquake. Coseismic slip of tens of metres can cause devastating shaking and tsunami, as most recently demonstrated by the moment magnitude (M_w) = 9 Tohoku, Japan, earthquake of 11 March 2011. The rupture and the subsequent evolution of stress and strain leading to the next earthquake is loosely called a subduction earthquake cycle (SEC). When the theory of plate tectonics had just gained wide acceptance, the role of mantle creep in SEC was postulated^{1,2} to reconcile the overall steady motion of tectonic plates and the jerky deformation at their boundaries due to great earthquakes. In this early model, SEC stress fluctuations propagate as one-dimensional diffusion waves through an elastic plate overlying a viscous asthenosphere (Fig. 1a). Its prediction that the strongest elastic deformation should be found near plate boundaries proved correct and provided guidance for geodetic studies of SEC in the ensuing decades. Maxwell viscoelasticity was first applied to earthquake models in the 1970s (ref. 3), in recognition that the deep Earth can elastically transmit stress yet behaves like a viscous fluid that hosts mantle convection and causes delayed response to the

removal of surface ice load (see review in ref. 4). Two-dimensional modelling efforts (Fig. 1b) in the 1980s (see review in ref. 5), especially those to explain century-long deformation records of the Nankai subduction zone, southwest Japan^{6,7}, established that three primary processes took place after each great earthquake (Fig. 2). These are: (1) continuing slip of the fault (now called ‘afterslip’), most evidently downdip of the rupture zone, (2) viscoelastic relaxation of the earthquake-induced stress, and (3) relocking of the subduction fault. These modelling efforts also led to the view that asthenosphere viscosity at subduction zones is about 10^{19} Pa s, one to two orders of magnitude lower than the global average⁸. However, the two-dimensional perspective failed to account for the fact that the duration of postseismic relaxation scales with rupture length in the third dimension. The consequent lack of universal success raised doubts about the importance of viscous deformation in SEC⁹.

The launch of the Global Positioning System (GPS) revolutionized crustal deformation measurements. Since the early 1990s, GPS measurements, either in campaign style or as continuous monitoring, have delineated patterns of co-, post- and interseismic deformation for many

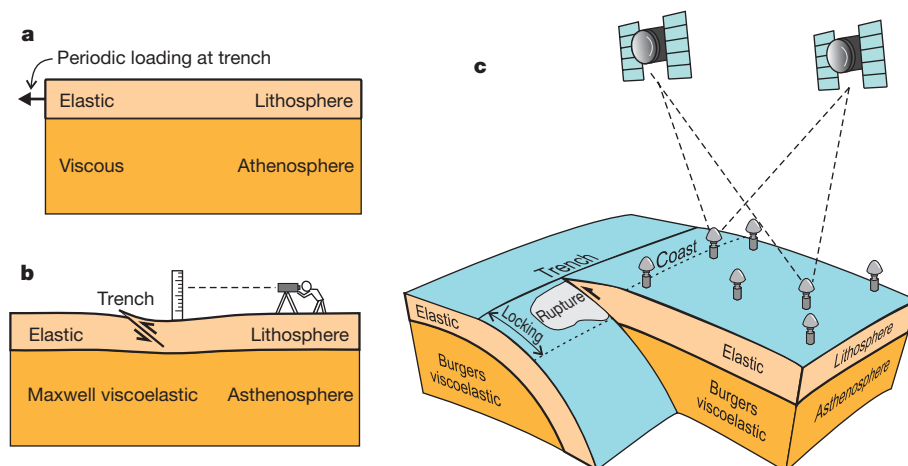


Figure 1 | Representative SEC models and development of observation technology. **a**, One-dimensional stress diffusion model of 1973 (ref. 2) based on ref. 1, a predictive model guiding future observations. **b**, Two-dimensional Maxwell viscoelastic model of 1984 (ref. 6). Observations were based mainly on

terrestrial geodesy, especially repeat levelling surveys. **c**, Three-dimensional Burgers viscoelastic model, used in this review. Today, space geodesy, especially GPS (satellites and antennas shown), is the most common means of observing contemporary crustal deformation.

¹Pacific Geoscience Centre, Natural Resources Canada, Geological Survey of Canada, Sidney, British Columbia V8L 4B2, Canada. ²School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3V6, Canada. [†]Present address: Geophysical Institute, University of Alaska, Fairbanks, Alaska 99775, USA.

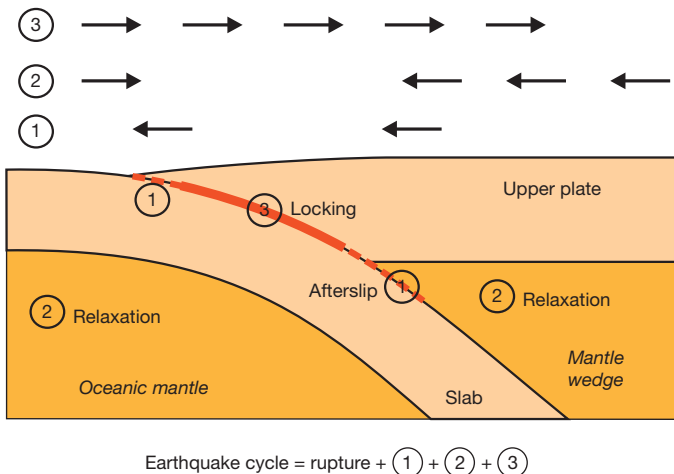


Figure 2 | Three primary processes after a subduction earthquake. (1) Aseismic afterslip occurs mostly around the rupture zone, (2) the coseismically stressed mantle undergoes viscoelastic relaxation, and (3) the fault is relocked. Arrows at the top show the sense of horizontal motion of Earth's surface, relative to distant parts of the upper plate, caused by each of these three processes.

subduction zones with unprecedented clarity. Developments in satellite radar interferometry and satellite gravimetry further improved the situation. In the meantime, progress in computing technology has enabled numerical viscoelastic SEC models that more accurately represent three-dimensional Earth structure and rupture geometry^{10–14} (Fig. 1c). However, the vast majority of the SEC models developed in this new era, especially those focused on fault locking or fault friction, assume a purely elastic Earth.

High-quality modern geodetic observations have been widely available for only one to two decades. For each subduction zone, they provide merely a 'snapshot' of its SEC evolution. The main reason for the popularity of the elastic model is the convenience it offers in inferring the apparent state of fault locking from each snapshot. However, it has become evident that these snapshots collectively reflect a common deformation cycle that includes the above-mentioned three primary processes (Fig. 2). Since the 1980s, the most important progress includes the stronger evidence for very low viscosities of the mantle wedge, the

recognition that postseismic viscoelastic relaxation consists of both transient and steady-state phases, as well as much improved knowledge of afterslip. In this review, we also emphasize that the characteristic time of the relaxation depends strongly on earthquake size.

Observing subduction earthquake cycles

When GPS first became available, it only confirmed what had already been inferred from sparse terrestrial geodetic measurements that many subduction faults are currently locked and accumulating strain energy for future earthquakes. An example is the Cascadia subduction zone^{15,16}, where an $M_w \approx 9$ subduction earthquake occurred in 1700, and GPS sites are now moving landward with respect to the remote areas of the upper plate (Fig. 3c). Soon, intriguing GPS velocities were reported from Chile^{17,18} and Alaska¹⁹ where a $M_w = 9.5$ and a $M_w = 9.2$ earthquake occurred in 1960 and 1964, respectively. These data show coastal sites to be moving landward as seen at other locked subduction zones, but inland sites some 200–400 km from the trench to be moving seaward (Fig. 3b). In 2004, a devastating $M_w = 9.2$ earthquake ruptured the Sumatra margin. Most strikingly, GPS sites in a very large area of the upper plate have been moving seaward ever since, in the same direction as their coseismic motion^{20–22} (Fig. 3a). In March 2011, the $M_w = 9$ Tohoku earthquake and its tsunami devastated northeastern Japan. Similarly to Sumatra, terrestrial GPS sites in northeastern Japan have been moving seaward since the earthquake (<http://www.gsi.go.jp/chibankansi/chikakukansi40005.html>; in Japanese).

A pattern of SEC evolution for giant earthquakes thus emerged. It features a progressive reversal of motion direction away from the rupture zone: wholesale seaward motion, followed by opposing motion of the coastal and inland areas, and eventual wholesale landward motion. This evolution, in a more convincing way than did the Nankai levelling data⁶, invalidates a popular belief that interseismic deformation is a subdued mirror image of coseismic deformation. In retrospect, we can see that the one-dimensional stress-diffusion model (Fig. 1a) four decades ago^{1,2} would have predicted the opposing motion currently seen at Chile and Alaska, but it took space geodesy to rediscover it.

Viscoelastic mantle relaxation

In laboratory experiments, rock creep exhibits an initial transient phase of rapid change and a subsequent steady-state phase²³ (see Box 1). Traditionally, the study of rock creep is focused on the steady-state

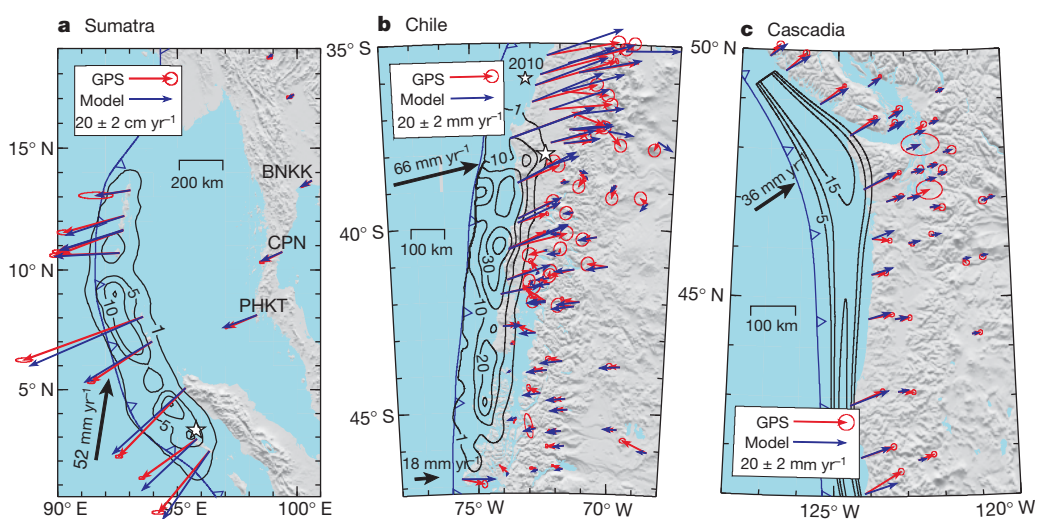


Figure 3 | GPS- (red) and model-predicted (blue) surface velocities for three subduction zones that are at different stages of the earthquake cycle. **a**, At Sumatra, one year after the $M_w = 9.2$ earthquake of 2004 (refs 20 and 21) (epicentre shown by star), all sites move seaward. Shown are ~ 1 -year average GPS velocities. More recent data show the same pattern²². Coseismic fault slip (contoured in metres) is based on ref. 56. Longer (~ 3 -years) time series from the three labelled far-field sites (BNKK, CPN, PHKT)³² helped constrain

afterslip and transient rheology (ref. 48). **b**, At Chile, four decades after the $M_w = 9.5$ earthquake of 1960, coastal and inland sites show opposing motion. Coseismic slip is from ref. 14. For sources of GPS data, see ref. 17. The northernmost areas show wholesale landward motion before the 2010 $M_w = 8.8$ Maule earthquake. **c**, At Cascadia, three centuries after the $M_w \approx 9$ earthquake of 1700, all sites move landward. The model is an updated version of ref. 8. A more comprehensive GPS compilation shows a similar deformation pattern¹⁶.

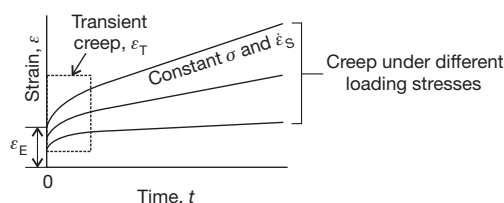
BOX 1

Steady-state and transient rheology

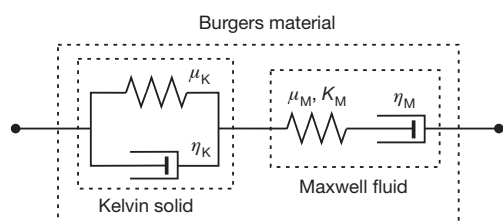
Under temperature (T) and pressure (P) conditions appropriate for the lower crust and mantle, if a rock specimen is loaded with a constant stress σ , it undergoes an initial elastic deformation ε_E , then transient creep of rapidly changing strain rate $\dot{\varepsilon}_T$, and eventually steady-state creep of constant strain rate $\dot{\varepsilon}_S$ (see Box 1 Fig. 1). Different loading stresses result in different steady-state strain rates. These pairs of σ and $\dot{\varepsilon}_S$ define the commonly seen steady-state flow law⁵⁷ $\dot{\varepsilon}_S = D\sigma^n \exp[-(Q + PV)/RT]$, where the parameter D depends on the rock composition, grain size, and fluid content, and n , Q , V and R are all constants.

The value of the stress exponent n depends on the microscopic deformation mechanism^{28,57}. For example, creep due to mass diffusion leads to $n = 1$, but creep due to migration of dislocations in crystalline lattice structure leads to $n \approx 3$. If transient creep is ignored, the viscoelastic rock deformation can be described using the Newtonian ($n = 1$) or non-Newtonian ($n > 1$) Maxwell fluid, which can be envisioned as an elastic element with shear modulus (rigidity) μ_M and bulk modulus K_M in series with this steady-state viscous element. For $n = 1$, the viscosity η_M is proportional to $D^{-1} \exp[(Q + PV)/RT]$, with the exact form depending on what stress component and strain-rate component are used for σ and $\dot{\varepsilon}_S$ and how the flow law is expressed in three dimensions. Most mantle convection, glacial isostatic adjustment, and SEC models assume $n = 1$, although some include strong temperature dependence and/or spatial variation. This viscosity is a simplified representation of a more complex system that may actually exhibit nonlinearity⁴. Its correspondence with the effective viscosity (a scaled ratio of stress to strain rate) of nonlinear flow cannot be quantified if the stress changes rapidly with time as in postseismic deformation.

A widely used parameterization of the transient creep is the Kelvin (or Kelvin–Voigt) solid, which can be envisioned as an elastic element of shear modulus μ_K in parallel with a viscous element of viscosity η_K . If the transient creep is considered, the simplest model to depict the viscoelastic rock deformation is the bi-viscous Burgers rheology²⁴, that is, a Kelvin solid in series with a Newtonian Maxwell fluid (see Box 1 Fig. 2). The spatial variation and stress-dependence of η_K are poorly known, although the viscosity ratio $\beta = \eta_M/\eta_K$ is sometimes assumed to be a constant. More complex creep behaviour can be described by including additional Kelvin and/or Maxwell elements. There is a recent attempt to modify the Burgers rheology to include non-Newtonian steady-state creep, but based on an assumption that a stress-independent β can still be defined³⁴.



Box 1 Fig. 1 | Strain evolution of a rock specimen under constant stress.



Box 1 Fig. 2 | Composition of the bi-viscous Burgers rheology.

behaviour that is appropriate for long-term processes such as mountain building and mantle convection. The viscous component of Maxwell viscoelasticity corresponds to the steady-state creep. The Maxwell rheology is found to be adequate in modelling century- to millennium-scale glacial isostatic adjustment^{24,25} and decadal-scale SEC deformation in Chile^{14,17,26} or Alaska^{12,13}.

The widely used elastic model and a steady-state mantle wedge viscosity of about 10^{19} Pa s in most Maxwell SEC models⁸ cannot be reconciled. What may support the elastic model is the enduring notion since 1935 (ref. 25) that upper-mantle viscosity is about 10^{21} Pa s, a value that would render a Maxwell mantle mostly elastic over an SEC of a few centuries. The upper-mantle viscosity structure has been much refined, but a global average of 10^{20} – 10^{21} Pa s is still valid²⁷. However, there is mounting evidence for lower viscosity at subduction zones, including laboratory measurements of relevant rocks²⁸, inference of small-scale convection in the backarc based on heat flow observations²⁹, constraints from topography and geoid anomalies³⁰, glacial isostatic adjustment analyses³¹, as well as the inferences from postseismic deformation⁸. The steady-state viscosity used to explain satellite gravity observations following the 2004 Sumatra earthquake is 10^{18} – 10^{19} Pa s (refs 32 and 33).

Continuous GPS monitoring suggests that immediate postseismic deformation and its change with time are much faster than the longer-term behaviour. From rock physics and experiments, it is logical to infer that the rapid short-term deformation reflects not only afterslip but also transient rock creep^{34,35}. In recent postseismic deformation models for Sumatra^{20,32,33} and the general SEC model in this review (Fig. 1c), a bi-viscous Burgers rheology is employed. The viscoelastic relaxation of the earthquake-induced stress thus has two timescales, a shorter one for the transient viscosity and a longer one for the steady-state viscosity (see Box 1).

Afterslip and relocking of the fault

The ubiquitous presence of afterslip is well recognized. A subduction fault is a zone of finite thickness with evolving internal structure³⁶, and its slip behaviour is affected by numerous factors including the roughness of the subducting seafloor³⁷ and amount of trench sediments³⁸. But to the first order it is regarded as a frictional contact. Afterslip is thus understood to be the velocity-strengthening behaviour of the fault areas around the rupture zone^{39,40}. Deeper than about 70–80 km, where the subduction interface is better described as a zone of localized viscous shear⁴¹, the distinction between afterslip (a fault process) and viscoelastic relaxation (a mantle process) becomes increasingly obscure.

A range of slip behaviour from stable sliding to unstable stick–slip can be explained using a rate-and-state friction law^{39,42}, but the application of the friction theory to realistic fault geometry and Earth rheology is still in its infancy. In most viscoelastic SEC models it is necessary to prescribe fault motion. The most common way of incorporating fault locking is to assign a backslip rate along the locked zone⁴³, assuming that steady subduction and its possible consequence in causing permanent surface deformation have been subtracted.

The characteristic timescale of afterslip, T_A , appears to be a few months to a few years⁴⁴. In some cases, if the Earth is assumed to be purely elastic, long-lasting afterslip is used to explain prolonged postseismic deformation⁴⁵. Conversely, there are also cases in which short-term postseismic deformation is explained using viscoelastic relaxation without resorting to afterslip^{20,46}. The truth lies in between, and determining T_A and the importance of afterslip relative to viscoelastic relaxation in controlling short-term post-seismic deformation remains a challenge.

In the three primary processes following an earthquake (Fig. 2), fault locking has the longest timescale (T_L), which is simply the length of the interseismic period. Over the past two decades, it has become a common mistake to equate ‘interseismic’ (a time concept) with ‘elastic deformation’ (a physical process). Even in recent three-dimensional viscoelastic models, the effect of fault locking is often simulated using an elastic Earth^{12–14}. Although the relaxation of the earthquake-induced stress is listed as one primary process in Fig. 2, it cannot be overemphasized that

the other two primary processes also involve their own viscoelastic relaxation^{6,26,43}. The stress built up by fault locking is partially relaxed at the same time, regardless of how long after the earthquake and whether the deformation pattern is still changing with time. Given the same locked zone and at a very long time after an earthquake, elastic shortening of the upper plate occurs in a broader zone if the mantle is undergoing viscoelastic relaxation but in a narrower zone if the mantle is purely elastic. For this reason, the maximum depth of locking is usually over-predicted if the viscous effect is not or insufficiently accounted for¹⁵, such as in all the locking models for northeastern Japan published before the Tohoku earthquake⁴⁷.

Piecing together an earthquake cycle

Using a numerical SEC model, we demonstrate that the drastically different deformation patterns at the three margins shown in Fig. 3 simply reflect different stages of a common evolution process. Although a site-specific three-dimensional model is needed to describe the structure and kinematics of each margin, the same rheological parameters are used for all the margins. In effect, there is only one physical model. The three margins differ from one another in many aspects, such as the age of the subducting plate and convergence rate. That a similar viscosity structure can be used to explain their first-order SEC is consistent with the observation that the arc-backarc regions of most subduction zones share a similar thermal state^{29,41}.

This model is an improvement over previously published viscoelastic models for these margins mentioned herein. For example, the use of the finite-element method allows the incorporation of realistic fault geometry and rupture distribution for Sumatra, and both afterslip and transient rheology are included for all the three margins. Because the transient rheology still has a small influence even a few decades after the earthquake, the steady-state viscosity required to explain the Chile observations is slightly lower than in previous models^{17,26}. Other important model features (Fig. 1c) include the spherical Earth geometry, which affects far-field deformation, and the presence of an elastic slab, which strongly controls the flow pattern of the mantle material and hence its relaxation process. A continuous distribution of afterslip is assigned that complements coseismic slip in the strike direction, tapers in the updip and downdip directions, and decays with time. Technical details of the modelling and the method of prescribing afterslip by trial-and-error are explained in ref. 48.

The relaxation time of a uniform Maxwell body of viscosity η_M and rigidity μ_M is $\tau_M = \eta_M/\mu_M$. For the subduction zone system, a mixture of elastic and viscoelastic materials, we define an empirical steady-state system-relaxation time $T_M = (M_o/M_o^0)\tau_M$, where τ_M is the Maxwell time of the isoviscous mantle wedge, M_o is the seismic moment (the product of rigidity, rupture area, and mean slip) of the earthquake that causes the stress perturbation, and M_o^0 is a reference moment. The relationship between M_o and M_w (ref. 49) gives $\log(M_o/M_o^0) = 1.5(M_w - M_w^0)$, where M_w^0 corresponds to M_o^0 . The role of rupture length, which two-dimensional models fail to account for, is reflected in this scaling factor. Similarly, we define an empirical transient system-relaxation time $T_K = (M_o/M_o^0)\tau_K$, where $\tau_K = \eta_K/\mu_K$ is the Kelvin time for the mantle wedge of transient viscosity η_K and rigidity μ_K . We assume $\mu_M = \mu_K = 64$ GPa, $\eta_M = 10^{19}$ Pa s, and $\eta_K = 5 \times 10^{17}$ Pa s. These viscosity values are similar to those used in previous Sumatra models^{20,32,33} that employed an analytical solution in a spherical viscoelastic Earth. For $M_w^0 = 8.4$, these parameters give $T_K \approx 4$ yr and $T_M \approx 80$ yr for an earthquake of $M_w = 9.2$. We also assume that the mantle wedge is less viscous than the oceanic mantle by a factor of ten because of the presence of fluids from the dehydrating slab^{8,28,30}. The higher-viscosity 10^{20} Pa s for the oceanic mantle is similar to the global mantle average²⁷. To focus on the first-order physics, we have intentionally ignored the spatial variations of the viscosities through their dependence on temperature T , pressure P and fluid contents. Uncertainties and approximations in the viscosity structure are responsible for the arbitrariness in the choice of M_o^0 in scaling T_M and T_K with earthquake size.

In explaining geodetic observations, there is a trade-off between T_A , T_K and T_M , reflecting uncertainties in the parameter values. η_M is better constrained by the observed opposing motion in Chile but can vary by a factor of two in the Sumatra and Cascadia models without significantly affecting the fit to the GPS data. Up to three years of GPS data from Sumatra require both afterslip and transient mantle rheology and help to determine T_A and η_K (ref. 48). New observations following more recent great earthquakes such as the $M_w = 8.8$ Maule event in 2010, just north of the 1960 Chile rupture (Fig. 3b), and the Tohoku event in 2011 will surely provide better constraints in the near future.

The three primary processes (Fig. 2) begin immediately after the earthquake, but they take turns to play dominant roles. At times comparable to T_A and T_K , the effect of Maxwell relaxation and fault locking are overshadowed by that of the afterslip and transient (Kelvin) rheology. The dominance of the afterslip and transient rheology cause rapid seaward motion at Sumatra (Fig. 3a). The Sumatra results also explain the present wholesale seaward motion in northeast Japan following the Tohoku earthquake. At times comparable to T_M , the effects of both Maxwell relaxation and locking become dominant, but the two processes give rise to contrasting surface motions (Fig. 2). For Chile (Fig. 3b), sites far away from the trench are still moving seaward because of Maxwell relaxation, but coastal sites are already moving landward because of fault locking. The Chile results also explain the opposing motion currently seen in Alaska^{12,13}. Eventually, at times much longer than T_M , the earthquake-induced stresses will be sufficiently relaxed, and the effect of fault locking dominates the entire upper plate. This is the situation currently seen at Cascadia (Fig. 3c). The Cascadia results also explain the wholesale landward motion observed before the Tohoku earthquake⁴⁷.

Scaling with earthquake size

The reason that T_K and T_M scale with the seismic moment is that the timescale of relaxation depends on the initial stress perturbation. For example, longer ruptures tend to excite longer-wavelength relaxation modes, which are generally associated with correspondingly longer relaxation times owing to the space-dispersive nature of the relaxation process⁵⁰. Therefore, after a smaller earthquake the effect of fault locking becomes dominant more quickly, such that the Sumatra-type seaward motion and the Chile-type opposing motion are short-lived or even absent²⁶. Failure to recognize this scaling was a fatal limitation of earlier two-dimensional models and caused doubts about the necessity of viscoelasticity. Here we illustrate the general role of earthquake size (Fig. 4), using idealized models with the same rheology as discussed above.

These models show how the surface of the upper plate progressively reverses its sense of motion from seaward to landward as the effect of fault locking becomes more dominant. The Extra-large model ($M_w \approx 9.2$), despite its simplicity, is consistent with predictions using models for the three giant events shown in Fig. 3 and current deformation in Alaska and northeastern Japan. It indicates that opposing motion will continue at Chile and Alaska for another few decades and will prevail at Sumatra and northeast Japan within a decade or two. For the Large model ($M_w \approx 8.8$), complete reversal from seaward to landward motion will take a decade or two. We predict this will happen in the area of the $M_w = 8.8$ Maule event of 2010, where wholesale landward motion was observed before the earthquake (northernmost part of Fig. 3b) but wholesale seaward motion is taking place at present⁵¹. Published examples for the quick reversal illustrated by the Medium model ($M_w \approx 8.4$) include two $M_w \approx 8$ –8.1 earthquakes in 1995 in Antofagasta, northern Chile⁵² and Colima-Jalisco, Mexico¹¹. Some of the GPS sites in the region of the $M_w = 8.3$ Kuril event in 2006 (ref. 46) have now reversed their motion. At the time of the $M_w = 8.4$ Arequipa, Peru, earthquake of 2001 (ref. 40), there was only one continuous GPS station (AREQ) in the affected area, some 200 km from the trench, and it reversed its motion about 6 to 7 years after the earthquake. The exact timing of the reversal depends also on the plate convergence rate, the amount of afterslip, and structural/geometrical details, and therefore may differ from these idealized models. The afterslip behaviour may vary from

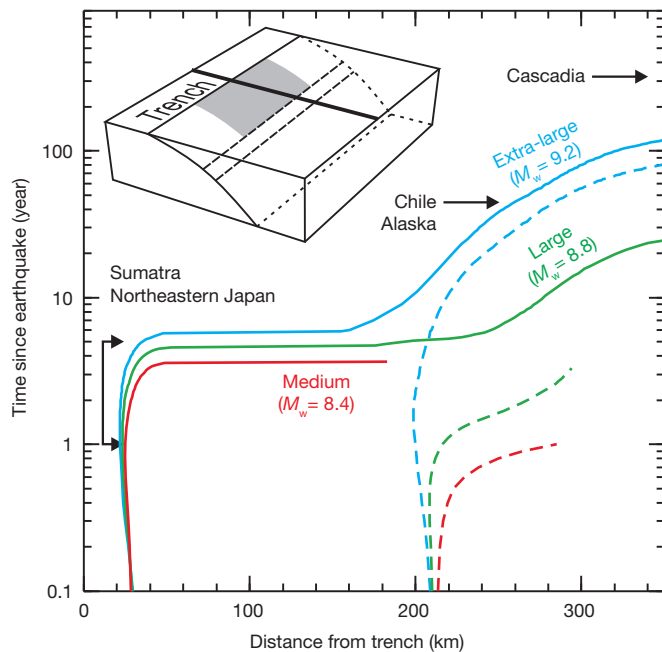


Figure 4 | Evolution of seaward-landward motion transition for different earthquake sizes. For each model, areas to the left of the curve move landward, and those to the right move seaward. The inset shows the model geometry and the line of symmetry (thick line) along which results are shown. Solid and dashed colour lines are results with and without afterslip, respectively. Within each trio, the models differ in rupture length, coseismic slip (assumed to scale with rupture length), and hence earthquake magnitude. The coseismic slip tapers both downdip and updip. The ensuing interseismic locking and its downdip transition (dashed lines) extend to the entire strike length, regardless of the rupture length. Except for the simple geometry and a convergence rate of 5 cm yr^{-1} , all parameters are the same as for Fig. 3. The Extra-large model qualitatively applies to today's deformation fields of the five subduction zones as indicated.

event to event depending on such factors as the structure and fluid pressure of the fault zone.

Also important, but not illustrated here, are along-strike variations in the degree of interseismic locking, as often inferred from geodetic measurements^{19,53}. Slowly slipping fault segments serve to load neighbouring, locked patches and thus affect the time and size of future rupture. Such inhomogeneous locking affects interseismic deformation pattern, and more so in the near field.

Looking into the future

Understanding the physics of interseismic deformation helps us to assess seismic and tsunami hazards, because it enables us to infer the locking state of the subduction fault. If viscous mantle creep is ignored, all surface deformation has to be explained by the locking and slip of various parts of the fault, leading to an incorrect depiction of the locking state. Fault-locking scenarios based on elastic models, either via forward modelling or inversion, should be reassessed to include the effect of viscoelastic relaxation. This also applies to other tectonic settings.

Although our present knowledge of Earth's viscoelasticity helps us to address the first-order physics of SEC, there are critical needs for better laboratory and field constraints on the transient rheology and a better understanding of the role of nonlinear rheology. The importance of better characterization and understanding of afterslip is illustrated in Fig. 4 by the large difference in short-term postseismic deformation with and without assuming afterslip. Measurements near the rupture zone are particularly useful. In this regard, ongoing or planned seafloor experiments in the rupture area of the 2011 Tohoku earthquake, such as GPS–acoustic surveys, pressure monitoring, and borehole strain measurements, promise breakthrough discoveries in the next few years.

Future seafloor geodesy can also fill the knowledge gap of SEC deformation seaward of the trench and yield critical constraints on the rheology of oceanic mantle.

The SEC deformation needs to be understood in a broad geodynamic context. It seems that the discovery of episodic slow slip events, sometimes accompanied with low-frequency seismic tremor, will soon lead to much improved understanding of the mechanics of the subduction fault and the SEC⁵⁴. Viscoelasticity gives rise to an asymmetry in SEC, that is, coseismic deformation is confined to be near the rupture zone, but interseismic deformation occurs in a broader region⁵⁵. The intriguing implications of this asymmetry for geodynamics and earthquake and tsunami hazards are yet to be explored.

1. Elsasser, W. M. in *The Application of Modern Physics to the Earth and Planetary Interiors* (ed. Runcorn, S. K.) 223–246 (John Wiley, 1969).
 2. Bott, M. H. P. & Dean, D. S. Stress diffusion from plate boundaries. *Nature* **243**, 339–341 (1973).
 3. Nur, A. & Mavko, G. Postseismic viscoelastic rebound. *Science* **183**, 204–206 (1974).
 4. Schubert, G., Turcotte, D. L. & Olsen, P. *Mantle Convection in the Earth and Planets* (Cambridge University Press, 2001).
- This book provides a comprehensive review of mantle rheology, Earth structure, theory of mantle convection, and fundamental questions to be addressed by the study of geodynamics.**
5. Cohen, S. Numerical models of crustal deformation in seismic zones. *Adv. Geophys.* **41**, 133–231 (1999).
 6. Thatcher, W. & Rundle, J. B. A viscoelastic coupling model for the cyclic deformation due to periodically repeated earthquakes at subduction zones. *J. Geophys. Res.* **89**, 7631–7640 (1984).
- This paper explains century-long levelling data from southwest Japan in terms of earthquake-cycle deformation and discusses short-term versus long-term postseismic deformation.**
7. Matsu'ura, M. & Sato, T. A dislocation model for the earthquake cycle at convergent plate boundaries. *Geophys. J. Int.* **96**, 23–32 (1989).
 8. Wang, K. in *The Seismogenic Zone of Subduction Thrust Faults* (eds Dixon, T. H. & Moore, J. C.) 540–574 (Columbia University Press, 2007).
 9. Savage, J. C. & Thatcher, W. Interseismic deformation at the Nankai Trough, Japan, subduction zone. *J. Geophys. Res.* **97**, 11117–11135 (1992).
 10. Wang, K., He, J., Dragert, H. & James, T. S. Three-dimensional viscoelastic interseismic deformation model for the Cascadia subduction zone. *Earth Planets Space* **53**, 295–306 (2001).
 11. Márquez Azúa, B., DeMets, C. & Masterlark, T. Strong interseismic coupling, fault afterslip, and viscoelastic flow before and after the Oct. 9, 1995 Jalisco-Colima earthquake: continuous GPS measurements from Colima, Mexico. *Geophys. Res. Lett.* **29**, 1281 (2002).
 12. Suito, H. & Freymueller, J. T. A viscoelastic and afterslip postseismic deformation model for the 1964 Alaska earthquake. *J. Geophys. Res.* **114**, B11404 (2009).
 13. Ali, S. T. & Freed, A. M. Contemporary deformation and stressing rates in Southern Alaska. *Geophys. J. Int.* **183**, 557–571 (2010).
 14. Moreno, M. S. et al. Heterogeneous plate locking in the South-Central Chile subduction zone: building up the next great earthquake. *Earth Planet. Sci. Lett.* **305**, 413–424 (2011).
 15. Wang, K., Wells, R., Mazzotti, S., Hyndman, R. D. & Sagiya, T. A revised dislocation model of interseismic deformation of the Cascadia subduction zone. *J. Geophys. Res.* **108** (B1), 2026 (2003).
 16. McCaffrey, R. et al. Fault locking, block rotation and crustal deformation in the Pacific Northwest. *Geophys. J. Int.* **169**, 1315–1340 (2007).
 17. Wang, K. et al. Crustal motion in the zone of the 1960 Chile earthquake: detangling earthquake-cycle deformation and forearc-sliver translation. *Geochim. Geophys. Geosyst.* **8**, Q10010 (2007).
 18. Khazaradze, G. et al. Prolonged post-seismic deformation of the 1960 great Chile earthquake and implications for mantle rheology. *Geophys. Res. Lett.* **29**, 2050 (2002).
 19. Freymueller, J. T. et al. in *Active Tectonics and Seismic Potential of Alaska* (eds Freymueller, J. T., Haeussler, P. J., Wesson, R. L. & Ekstrom, G.) *Geophys. Monogr. Ser.* 179 1–42 (American Geophysical Union, 2008).
 20. Pollitz, F., Banerjee, P., Grijalva, K., Nagarajan, B. & Burgmann, R. Effect of 3-D viscoelastic structure on post-seismic relaxation from the 2004 $M = 9.2$ Sumatra earthquake. *Geophys. J. Int.* **173**, 189–204 (2008).
 21. Shearer, P. & Bürgmann, R. Lessons learned from the 2004 Sumatra-Andaman megathrust rupture. *Annu. Rev. Earth Planet. Sci.* **38**, 103–131 (2010).
- This review of the study of the 2004 Sumatra earthquake shows how geology, seismology, and space geodesy can be combined to investigate coseismic and postseismic deformation and tsunami processes.**
22. Grijalva, K. A., Bürgmann, R. & Banerjee, P. Using postseismic geodetic data to constrain the Sunda downdip transition zone. *Eos (Fall Meet. Suppl.)* **90** (52), abstr. T13E-01 (2009).
 23. Weertman, J. & Weertman, J. R. High temperature creep of rock and mantle viscosity. *Annu. Rev. Earth Planet. Sci.* **3**, 293–315 (1975).
 24. Peltier, W. R., Wu, P. & Yuen, D. A. in *Anelasticity in the Earth* (eds Stacey, F. D., Paterson, M. S. & Nicolas, A.) *Geodynamics Ser.* 4 59–77 (American Geophysical Union, 1981).
 25. Mitrovica, J. X. Haskell [1935] revisited. *J. Geophys. Res.* **101** (B1), 555–569 (1996).

26. Hu, Y., Wang, K., He, J., Klotz, J. & Khazaradze, G. Three-dimensional viscoelastic finite element model for post-seismic deformation of the great 1960 Chile earthquake. *J. Geophys. Res.* **109**, B12403 (2004).
27. Moucha, R., Forte, A. M., Mitrova, J. X. & Daradich, A. Lateral variations in mantle rheology: implications for convection related surface observables and inferred viscosity models. *Geophys. J. Int.* **169**, 113–135 (2007).
28. Hirth, G. & Kohlstedt, D. L. in *Inside the Subduction Factory* (ed. Eiler, J.) 83–105 (American Geophysical Union, 2003).
29. Currie, C. A. & Hyndman, R. D. The thermal structure of subduction zone back arcs. *J. Geophys. Res.* **111**, B08404 (2006).
30. Billen, M. I. & Gurnis, M. A low viscosity wedge in subduction zones. *Earth Planet. Sci. Lett.* **193**, 227–236 (2001).
31. James, T. S., Clague, J. J., Wang, K. & Hutchinson, I. I. Postglacial rebound at the northern Cascadia subduction zone. *Quat. Sci. Rev.* **19**, 1527–1541 (2000).
32. Panet, I. *et al.* Upper mantle rheology from GRACE and GPS postseismic deformation after the 2004 Sumatra–Andaman earthquake. *Geochem. Geophys. Geosyst.* **11**, Q06008 (2010).
33. Han, S.-C., Sauber, J., Luthcke, S. B., Ji, C. & Pollitz, F. F. Implications of postseismic gravity change following the great 2004 Sumatra–Andaman earthquake from the regional harmonic analysis of GRACE intersatellite tracking data. *J. Geophys. Res.* **113**, B11413 (2008).
34. Freed, A. M., Hirth, G. & Behn, M. D. Using short-term postseismic displacements to infer the ambient deformation conditions for the upper mantle. *J. Geophys. Res.* **117**, B01409 (2012).
35. Pollitz, F. F. Transient rheology of the uppermost mantle beneath the Mojave Desert, California. *Earth Planet. Sci. Lett.* **215**, 89–104 (2003).
36. Bachmann, R. *et al.* Exposed plate interface in the European Alps reveals fabric styles and gradients related to an ancient seismogenic coupling zone. *J. Geophys. Res.* **114**, B05402 (2009).
37. Wang, K. & Bilek, S. L. Do subducting seamounts generate or stop large earthquakes? *Geology* **39**, 819–822 (2011).
38. Scholl, D. W., Kirby, S. H. & von Huene, R. Exploring a link between great and giant megathrust earthquakes and relative thickness of sediment and eroded debris in the subduction channel to roughness of subducted relief. *AGU Fall Meet. abstr.* T14B-01 (2011).
39. Scholz, C. H. Earthquakes and friction laws. *Nature* **391**, 37–42 (1998).
This paper reviews fundamentals of rate- and state-dependent friction and applications to earthquakes and aseismic fault slip.
40. Perfettini, H., Avouac, J.-P. & Ruegg, J.-C. Geodetic displacements and aftershocks following the 2001 $M_w = 8.4$ Peru earthquake: implications for the mechanics of the earthquake cycle along subduction zones. *J. Geophys. Res.* **110**, B09404 (2005).
41. Wada, I. & Wang, K. Common depth of slab-mantle decoupling: reconciling diversity and uniformity of subduction zones. *Geochem. Geophys. Geosyst.* **10**, Q10009 (2009).
42. Hetland, E. A. & Simons, M. Post-seismic and interseismic fault creep. II: Transient creep and interseismic stress shadows on megathrusts. *Geophys. J. Int.* **181**, 99–112 (2010).
43. Savage, J. C. A dislocation model of strain accumulation and release at subduction zones. *J. Geophys. Res.* **88**, 4984–4996 (1983).
44. Pritchard, M. E. & Simons, M. An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behaviour. *J. Geophys. Res.* **111**, B08405 (2006).
45. Sawai, Y. *et al.* Transient uplift after a 17th-century earthquake along the Kuril subduction zone. *Science* **306**, 1918–1920 (2004).
46. Kogan, M. G. *et al.* The mechanism of postseismic deformation triggered by the 2006–2007 great Kuril earthquakes. *Geophys. Res. Lett.* **38**, L06304 (2011).
47. Simons, M. *et al.* The 2011 magnitude 9.0 Tohoku-oki earthquake: mosaicking the megathrust from seconds to centuries. *Science* **332**, 1421–1425 (2011).
48. Hu, Y., & Wang, K. Spherical-Earth finite element model of short-term postseismic deformation following the 2004 Sumatra earthquake. *J. Geophys. Res.* (in the press).
49. Hanks, T. C. & Kanamori, H. A moment magnitude scale. *J. Geophys. Res.* **84**, 2348–2350 (1979).
50. Segall, P. *Earthquake and Volcano Deformation* (Princeton University Press, 2010).
51. Brooks, B. A. *et al.* Andean backarc deformation and the plate boundary earthquake cycle. *AGU Fall Meet. abstr.* S14A-08 (2011).
52. Khazaradze, G. & Klotz, J. Short and long-term effects of GPS measured crustal deformation rates along the south central Andes. *J. Geophys. Res.* **108** (B6), 2289 (2003).
53. Wallace, L. M. & Beavan, J. Diverse slow slip behavior at the Hikurangi subduction margin, New Zealand. *J. Geophys. Res.* **115**, B12402 (2010).
54. Peng, Z. & Gornberg, J. An integrated perspective of the continuum between earthquakes and slow slip phenomena. *Nature Geosci.* **3**, 599–607 (2010).
55. Wang, K. Coupling of tectonic loading and earthquake fault slips at subduction zones. *Pure Appl. Geophys.* **145**, 537–559 (1995).
56. Chlieh, M. *et al.* Coseismic slip and afterslip of the great M_w 9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97** (1A), S152–S173 (2007).
57. Bürgmann, R. & Dresen, G. Rheology of the lower crust and upper mantle: evidence from rock mechanics, geodesy and field observations. *Annu. Rev. Earth Planet. Sci.* **36**, 531–567 (2008).

Acknowledgements M. Kogan and M. Chlieh provided unpublished information on GPS station reversal after the 2006 Kuril and 2001 Peru earthquakes, respectively. This is Geological Survey of Canada contribution 20110422.

Author Contributions K.W. designed the study and prepared the manuscript. Y.H. did the numerical modelling. J.H. wrote the modelling code and contributed to the modelling.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to K.W. (kwang@nrcan.gc.ca).

A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism

Mark A. Herman¹, Odile D. Peroni¹, Jorge Villoria¹, Michael R. Schön², Nada A. Abumrad³, Matthias Blüher⁴, Samuel Klein³ & Barbara B. Kahn¹

The prevalence of obesity and type 2 diabetes is increasing worldwide and threatens to shorten lifespan. Impaired insulin action in peripheral tissues is a major pathogenic factor. Insulin stimulates glucose uptake in adipose tissue through the GLUT4 (also known as SLC2A4) glucose transporter, and alterations in adipose tissue GLUT4 expression or function regulate systemic insulin sensitivity. Downregulation of human and mouse adipose tissue GLUT4 occurs early in diabetes development. Here we report that adipose tissue GLUT4 regulates the expression of carbohydrate-responsive-element-binding protein (ChREBP; also known as MLXIPL), a transcriptional regulator of lipogenic and glycolytic genes. Furthermore, adipose ChREBP is a major determinant of adipose tissue fatty acid synthesis and systemic insulin sensitivity. We find a new mechanism for glucose regulation of ChREBP: glucose-mediated activation of the canonical ChREBP isoform (ChREBP- α) induces expression of a novel, potent isoform (ChREBP- β) that is transcribed from an alternative promoter. *ChREBP- β* expression in human adipose tissue predicts insulin sensitivity, indicating that it may be an effective target for treating diabetes.

Insulin resistance is a common complication of obesity and a major factor in the pathogenesis of type 2 diabetes and cardiovascular disease¹. Adipose tissue contributes to the development of obesity-related insulin resistance through increased release of fatty acids, altered adipokine secretion, and/or macrophage infiltration and cytokine release^{2,3}. Altered adipose tissue glucose metabolism is also an important cause of insulin resistance, and adipose tissue GLUT4, the major insulin-responsive glucose transporter, has a central role in systemic glucose metabolism^{1,4,5}. In insulin-resistant states, GLUT4 is downregulated in adipose tissue, but not in muscle¹, the major site of insulin-stimulated glucose uptake. In addition, mice with adipose-specific GLUT4 overexpression (AG4OX) have improved glucose homeostasis⁵ whereas adipose-specific GLUT4 knockout mice (AG4KO) have insulin resistance and type 2 diabetes⁴. We investigated how altering adipose tissue glucose flux regulates glucose homeostasis. We show that ChREBP, a glucose-responsive transcription factor that regulates fatty acid synthesis and glycolysis⁶, is highly regulated by GLUT4 in adipose tissue and is a key determinant of systemic insulin sensitivity and glucose homeostasis. Also, ChREBP in adipose tissue is required for the improved glucose homeostasis resulting from increased adipose GLUT4 expression. GLUT4-mediated glucose uptake induces ChREBP, which activates adipose tissue *de novo* lipogenesis (DNL). The latter is associated with enhanced insulin sensitivity^{7–10}. In obese humans, adipose *ChREBP* gene expression correlates with insulin sensitivity, suggesting that ChREBP protects against obesity-associated insulin resistance. In addition, we discovered a novel mechanism for glucose-regulated *ChREBP* expression involving a new isoform, *ChREBP- β* , which is expressed from an alternative promoter in a glucose- and ChREBP-dependent manner. In contrast, expression of the canonical *ChREBP- α* isoform is not regulated by glucose flux. However, glucose-induced ChREBP- α transcriptional activity increases *ChREBP- β* expression. Furthermore, expression of *ChREBP- β* is more highly regulated than *ChREBP- α* in adipose tissue in insulin-resistant states. Thus, activation of adipose

ChREBP, and particularly ChREBP- β , may be a novel strategy for preventing and treating obesity-related metabolic dysfunction and type 2 diabetes (Supplementary Fig. 1).

Glucose regulates adipose tissue ChREBP

To understand the mechanisms by which adipocytes respond to changes in glucose flux, we analysed global gene expression in adipose tissue from AG4OX and AG4KO mice. Gene-set enrichment analysis¹¹ demonstrated coordinate upregulation of DNL enzymes in AG4OX mice (Supplementary Table 1), which we confirmed (Fig. 1a). DNL enzymes were downregulated in AG4KO mice (Fig. 1a). Therefore we investigated the expression of the two major transcription factors known to regulate DNL enzymes, sterol regulatory-element-binding-protein 1c (*SREBP-1c*; also known as isoform 1c of *SREBF1*) and *ChREBP*^{6,12}. *ChREBP*, but not *SREBP-1c*, expression is increased 50% in AG4OX and decreased 44% in AG4KO adipose tissue compared to controls (Fig. 1b). *SREBP-1c* transcriptional activity is primarily determined by the accumulation of mature *SREBP-1c* in the nucleus¹³. However, the nuclear abundance of *SREBP-1c* is not increased in AG4OX adipose tissue (Supplementary Fig. 2a). Liver X receptors (LXR- α and LXR- β ; also known as NR1H3 and NR1H2, respectively) can regulate the expression of both *ChREBP* and *SREBP-1c*^{14,15}, and DNL enzymes¹⁶. Expression of canonical LXR targets do not change (Supplementary Fig. 2b) in AG4OX or AG4KO adipose tissue, indicating that LXR activity is unaffected and is not driving the changes in *ChREBP* or DNL enzyme expression. In contrast, expression of *Rgs16* and *Txnip*, two ChREBP transcriptional targets^{17–19} not known to be regulated by other lipogenic transcription factors, were reciprocally regulated in AG4OX and AG4KO mice (Supplementary Fig. 2c). In AG4KO and control mice (Fig. 1c), and also in 30 different mouse strains (Supplementary Fig. 3), adipose *ChREBP* expression strongly correlates with *GLUT4* expression. Expression of *ChREBP* transcriptional targets *FAS* (also known as *Fasn*) and *Acc1* (also known as

¹Division of Endocrinology, Diabetes and Metabolism, Beth Israel Deaconess Medical Center and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Städtisches Klinikum Karlsruhe, Clinic of Visceral Surgery, Karlsruhe 76133, Germany. ³Center for Human Nutrition, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴Department of Medicine, University of Leipzig, Leipzig 04103, Germany.

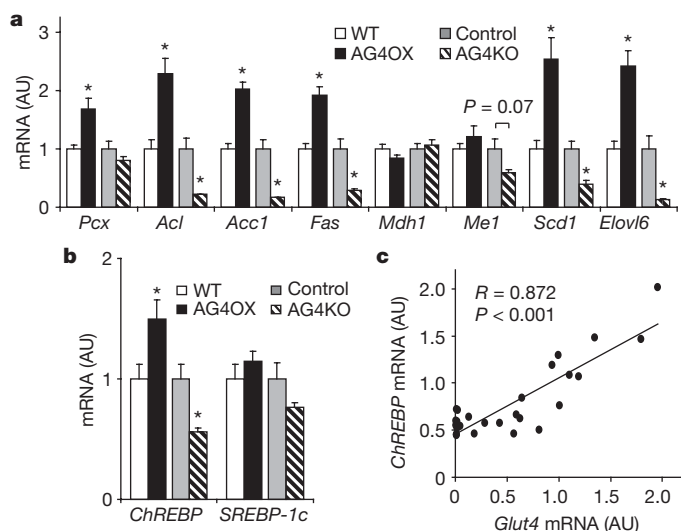


Figure 1 | Genetically altering adipose tissue glucose flux regulates the expression of *ChREBP* and its lipogenic targets. **a**, **b**, mRNA expression of fatty acid synthetic enzymes (**a**) and lipogenic transcription factors (**b**) in perigonadal fat from 6-week-old female mice ($n = 10$ –14 per group). AU, arbitrary units. WT, wild type. $*P < 0.05$ compared to respective controls. **c**, *ChREBP* and *Glut4* mRNA correlate highly in perigonadal WAT from control and AG4KO mice ($n = 27$). Values are means \pm standard error.

Acaca) also strongly correlates with *ChREBP* expression across these strains (Supplementary Fig. 3). Thus, adipose *ChREBP* may mediate the effects of altered GLUT4 expression and glucose flux on lipogenic enzyme expression.

ChREBP mediates effects of adipose GLUT4

Increased DNL in adipose tissue is associated with enhanced glucose tolerance and insulin sensitivity^{20,21}. Consistent with the increases in

ChREBP and DNL enzyme expression (Fig. 1a, b), DNL *in vivo* is increased in perigonadal and subcutaneous white adipose tissue (WAT) in AG4OX mice compared to wild type (Fig. 2a).

To determine whether adipose *ChREBP* is required for the increased adipose tissue DNL resulting from GLUT4 overexpression, we crossbred AG4OX mice with whole-body *ChREBP* KO mice⁶. DNL normalizes in adipose tissue after deletion of *ChREBP* in AG4OX mice (Fig. 2a). Glucose incorporation into newly synthesized fatty acids is also decreased in wild-type and AG4OX adipose tissue after *ChREBP* knockout (Supplementary Fig. 4). In liver, DNL is normal in AG4OX and AG4OX/*ChREBP* KO mice and is not decreased in *ChREBP* KO mice (Fig. 2a) contrasting with a prior report⁶. This difference may be attributable to differences in age, sex, genetic background and/or diet.

In parallel with a reduction in DNL in AG4OX adipose tissue after *ChREBP* deletion, DNL enzyme messenger RNA and protein expression levels also normalize (Fig. 2b and Supplementary Fig. 5). Thus, *ChREBP* is a major regulator of DNL enzymes in adipose tissue and mediates the changes in DNL enzyme expression resulting from increased glucose flux.

AG4OX mice are obese (Fig. 2c, d) but have enhanced glucose tolerance (Fig. 2e)⁵. Increased body weight (Fig. 2c), adiposity (Fig. 2d) and a tendency towards increased food intake normalize with *ChREBP* deletion; energy expenditure was normal (Supplementary Fig. 6). In spite of obesity, AG4OX mice are more insulin sensitive than wild-type mice (Supplementary Table 2). Decreased fed glycaemia (Supplementary Table 2), markedly enhanced glucose tolerance (Fig. 2e), and lower glycaemia during insulin tolerance testing (Fig. 2f) in AG4OX mice normalized after *ChREBP* deletion. Genetically ablating *ChREBP* in wild-type and AG4OX mice impairs insulin sensitivity (Supplementary Table 2). *ChREBP* KO mice show modest glucose intolerance (Fig. 2e), as reported⁶. Glycaemia in *ChREBP* KO mice remains elevated after food removal compared to wild-type mice (Fig. 2g). *ChREBP* knockout prevents the hypoglycaemia that develops

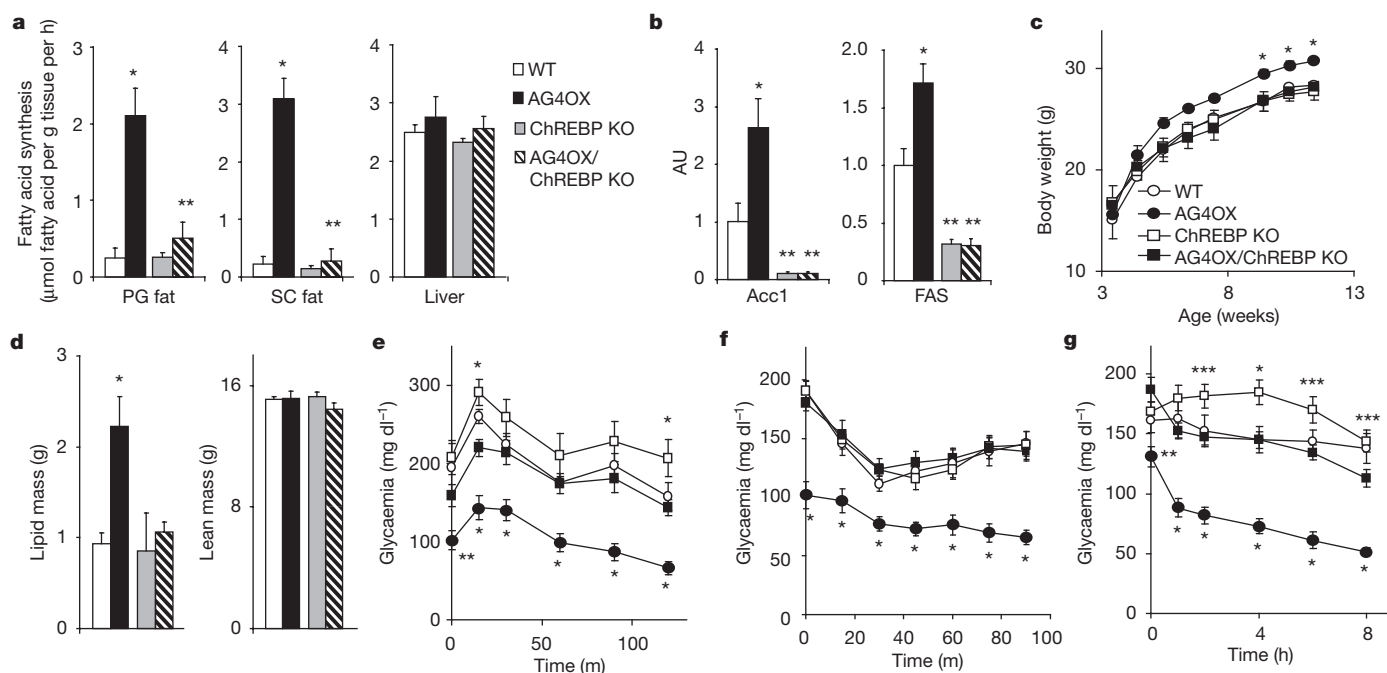


Figure 2 | *ChREBP* is essential for the effects of adipose tissue GLUT4 on adiposity, DNL, and glucose homeostasis. **a**, DNL measured *in vivo* in fed, 4-month-old male mice. PG, perigonadal; SC, subcutaneous. $n = 5$ –6 per group. **b**, Quantification of western blots of FAS and ACC1 in subcutaneous fat from fed, 6-month-old females ($n = 9$ –10 per group). **a**, **b**, $*P < 0.05$ versus same *ChREBP* genotype, different AG4OX genotype. $**P < 0.05$ versus same AG4OX genotype, different *ChREBP* genotype. **c**, Body weights in male mice

on chow ($n = 5$ –7 per group). $*P < 0.05$ versus all other groups at the indicated time. **d**, Body composition in 8-week-old, female mice (for **d**–**g**; $n = 10$ –12 per group). $*P < 0.01$ versus all others. **e**, **f**, Glucose tolerance test (**e**) and insulin tolerance test (**f**). $*P < 0.05$ versus all others; $**P < 0.05$ versus wild type and *ChREBP* KO. **g**, Glycaemia following food removal. $*P < 0.05$ versus all others; $**P < 0.05$ versus *ChREBP* KO and AG4OX/*ChREBP* KO; $***P < 0.05$ versus AG4OX and AG4OX/*ChREBP* KO. Values are means \pm standard error.

in AG4OX mice after food removal (Fig. 2g). ChREBP deletion did not decrease GLUT4 overexpression or glucose uptake (Supplementary Fig. 7) in AG4OX adipocytes, indicating that the mechanism by which ChREBP regulates glucose homeostasis is not by altering adipose tissue glucose uptake. The decreased insulin sensitivity and glucose tolerance in AG4OX/ChREBP KO mice is not due to elevated serum triglycerides or non-esterified fatty acids, as both metabolites decrease compared to AG4OX mice (Supplementary Table 2). The effects of ChREBP KO on metabolism and gene expression in AG4OX mice is likely to result from the absence of ChREBP specifically in adipose tissue, because *ChREBP* expression is increased in adipose tissue (Fig. 1b) but not in liver of AG4OX mice (Supplementary Fig. 8). Reversal of enhanced glucose tolerance and insulin sensitivity with normalization of DNL is

consistent with a role for adipose tissue DNL in improving systemic glucose metabolism.

Obesity regulates adipose tissue ChREBP

High-fat diet (HFD) causes obesity and insulin resistance and down-regulates GLUT4 expression selectively in adipose tissue¹. To investigate whether adipose ChREBP expression has a role in the insulin resistance resulting from adipose GLUT4 downregulation, we subjected mice to HFD. On chow, AG4OX mice are obese (Figs 2c, d and 3a). AG4OX mice do not gain more weight on HFD and their degree of obesity on both diets is similar to HFD-fed wild-type mice (Fig. 3a and Supplementary Fig. 9). HFD induces insulin resistance in both genotypes (Supplementary Table 3). In wild-type mice, HFD causes a diabetic glucose tolerance test (Fig. 3b) but not in AG4OX mice. The enhanced glucose tolerance in AG4OX mice does not result from changes in serum metabolites typically associated with insulin sensitivity. AG4OX mice have higher non-esterified fatty acids and triglycerides, and lower leptin levels on both diets (Supplementary Table 3), and lower adiponectin levels²² on chow compared to wild type. This pattern is thought to contribute to insulin resistance^{2,3,23} and not to enhanced insulin sensitivity.

HFD reduces expression of adipose tissue DNL enzymes^{24,25} which correlates with insulin resistance^{7–10}. Therefore, we examined whether downregulation of adipose *ChREBP* and DNL contributes to HFD-induced insulin resistance. HFD in wild-type mice reduces DNL in subcutaneous but not perigonadal adipose tissue, (Fig. 3c). In AG4OX, HFD markedly reduces DNL in WAT, but DNL remains 2–7-fold higher than in wild-type mice on HFD. This intermediate rate of DNL parallels the change in glucose tolerance in AG4OX mice on HFD (Fig. 3b). Changes in glucose incorporation into newly synthesized fatty acids (Supplementary Fig. 10) are similar to the changes in total (from all substrates) DNL (Fig. 3c) in WAT. In liver, DNL is not higher in AG4OX on either diet (Fig. 3c). HFD inhibits adipose DNL independently of adipose tissue glucose uptake in AG4OX mice, as GLUT4 protein levels (Supplementary Fig. 11) and glucose transport remain elevated²⁶.

To investigate the mechanisms underlying HFD-induced changes in adipose DNL, we examined the expression of lipogenic transcription factors and DNL enzymes. In wild type mice, HFD diminishes *ChREBP* but not *SREBP-1c* expression in WAT (Fig. 3d), as reported^{25,27}. In AG4OX WAT, HFD markedly decreases *ChREBP* and also *SREBP-1c* expression but *ChREBP* remains elevated compared to wild-type mice on HFD. Nevertheless, HFD reduces the degree of *ChREBP* induction in AG4OX adipose tissue despite persistent GLUT4 overexpression (Supplementary Fig. 11). Thus, HFD may regulate *ChREBP* expression in part independently of GLUT4 expression and glucose flux, consistent with HFD effects on DNL (Fig. 3c).

Many DNL enzymes that are upregulated in WAT of AG4OX on chow are downregulated in HFD-fed wild-type and AG4OX mice

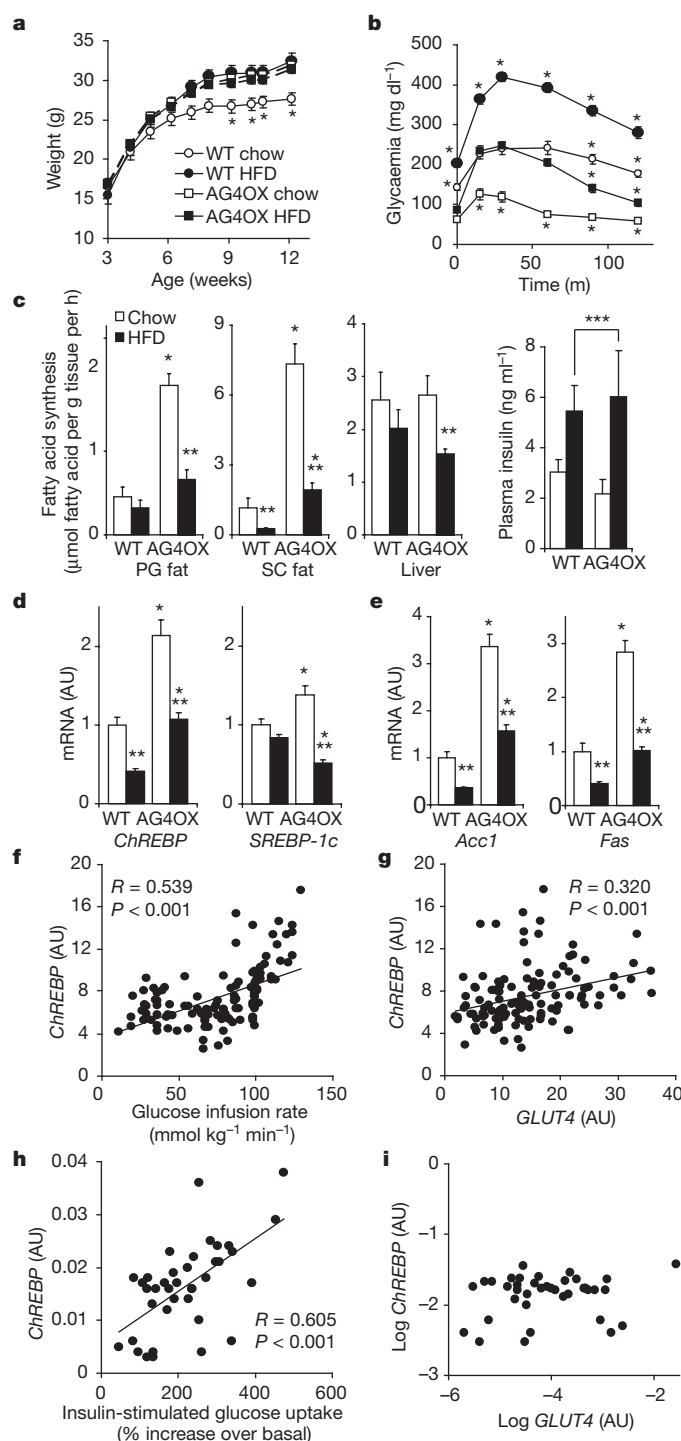


Figure 3 | ChREBP is regulated in mouse and human adipose tissue in pathological conditions. **a**, Body weight in chow- versus HFD-fed wild-type and AG4OX mice (9-week-old, $n = 8$ per group), $*P < 0.05$ chow-fed wild type compared to all others at the same time point. **b**, Glucose tolerance test in 7-week-old males ($n = 8$ per group). $*P < 0.001$ versus all others. **c**, Fatty acid synthesis measured *in vivo* in 5–6-month-old males ($n = 7$ per group). PG, perigonadal; SC, subcutaneous. Right, plasma insulin: HFD increases insulin ($***P = 0.01$ by two-way ANOVA). **d**, **e**, mRNA expression of *ChREBP* and *SREBP-1c* (**d**) and *Acc1* and *Fas* (**e**) in subcutaneous fat from 4-month-old male mice ($n = 8–14$ per group). For panels **c–e**, $*P < 0.05$ compared to same diet, different genotype. $**P < 0.05$ compared to same genotype, different diet. **f**, Subcutaneous fat *ChREBP* mRNA correlates with insulin sensitivity in non-diabetic, normal glucose-tolerant humans ($n = 123$). **g**, *ChREBP* and *Glut4* mRNA expression in subcutaneous fat correlate in this group. **h**, Subcutaneous fat *ChREBP* mRNA correlates highly with insulin sensitivity in obese, non-diabetic, BMI-matched humans ($n = 38$). **i**, *ChREBP* and *Glut4* mRNA expression do not correlate in this group. Values are means \pm standard error.

(Fig. 3e and Supplementary Fig. 12) but remain higher in AG4OX WAT compared to wild type in a pattern paralleling *ChREBP* expression (Fig. 3d) and DNL (Fig. 3c). In wild-type WAT, *ChREBP* downregulation probably accounts for HFD-induced downregulation of DNL gene expression, as WAT *SREBP-1c* expression is not altered. Furthermore, whole-body genetic ablation of *SREBP-1c* does not diminish expression of these genes in WAT²⁸.

In contrast, both *ChREBP* and *SREBP-1c* expression are downregulated in WAT in AG4OX mice on HFD compared to chow (Fig. 3d). Persistent modest elevation of *ChREBP* in WAT of AG4OX on HFD compared to wild type on HFD (Fig. 3d) is probably responsible for the increased DNL that contributes to improved glucose tolerance in AG4OX mice (Fig. 3b). Unlike in wild-type WAT, in liver, HFD does not downregulate *ChREBP* and DNL enzyme expression (Supplementary Fig. 13).

LXRs are unlikely to contribute to downregulation of *ChREBP* expression, DNL, or DNL enzyme expression in wild-type mice on HFD, because expression of LXRs and canonical LXR targets are either unchanged or modestly increased (Supplementary Fig. 14). MAX-like protein X (Mlx) is an obligate dimerization partner for *ChREBP* transcriptional activity²⁹. Neither HFD nor GLUT4 overexpression alters *MLX* expression in WAT (Supplementary Fig. 14).

We next sought to determine whether adipose *ChREBP* might contribute to regulating insulin sensitivity and glucose homeostasis in humans. In 123 non-diabetic individuals with normal glucose tolerance and widely ranging body mass index (BMI) values (Supplementary Table 4), *ChREBP* expression in subcutaneous WAT correlates strongly with insulin sensitivity measured during a euglycaemic and hyperinsulinaemic clamp procedure (Fig. 3f). Adipose tissue *ChREBP* correlates with *GLUT4* (Fig. 3g), consistent with a role for *ChREBP* in the beneficial effects of adipose tissue GLUT4 expression on glucose homeostasis.

Most, but not all, obese people are insulin resistant. To determine whether adipose *ChREBP* expression could have a role in regulating insulin sensitivity in obese people, we also investigated adipose

ChREBP expression in non-diabetic obese individuals with widely ranging insulin sensitivity (Supplementary Table 4). Adipose *ChREBP* expression was directly associated with insulin-stimulated glucose uptake during a clamp, independent of BMI (Fig. 3h). Surprisingly, *ChREBP* and *GLUT4* expression did not correlate in these obese individuals (Fig. 3i). Thus, adipose *ChREBP* expression may have beneficial effects on insulin sensitivity, even among obese subjects, and this can be independent of *GLUT4* expression.

Identification of a novel *ChREBP* isoform

The mechanism by which glucose activates *ChREBP* transcriptional activity is complex and controversial. It has been attributed to xylulose-5-phosphate-induced dephosphorylation of specific *ChREBP* residues³⁰. Alternatively, glucose or a metabolite may prevent the interaction of an amino-terminal inhibitory domain with a glucose-activating domain within the *ChREBP* protein³¹. Glucose and other carbohydrates can also induce *ChREBP* expression²⁷ but whether this contributes to *ChREBP* activity is unknown. The mechanism for carbohydrate-mediated *ChREBP* induction may involve 'feed-forward' auto-regulation, because a dominant-negative *MLX* mutant that inhibits *ChREBP* activity reduces endogenous *ChREBP* expression^{19,32}.

To investigate the molecular mechanisms by which *ChREBP* might transactivate its own expression, we searched the genomic sequence around the *ChREBP* transcriptional start site for a carbohydrate response element (ChoRE), which is defined by two E-boxes (CACGTG) separated by five nucleotides³³. We identified a single candidate ChoRE 17 kb upstream of the mouse *ChREBP* transcriptional start site (Supplementary Figs 15 and 16). Additionally, a separate E-box was identified 255 base pairs proximal to the putative ChoRE (Supplementary Fig. 16). In this region, the genomic sequence including the ChoRE and 5' E-box is highly conserved (Supplementary Figs 16 and 17). H3K4me3 and H3K4me1 histone methylation marks aligned with this conserved genomic region (Supplementary Fig. 18), suggesting the presence of an alternative promoter³⁴ and potentially an alternative first exon. By 5' rapid amplification of cDNA ends, we confirmed an

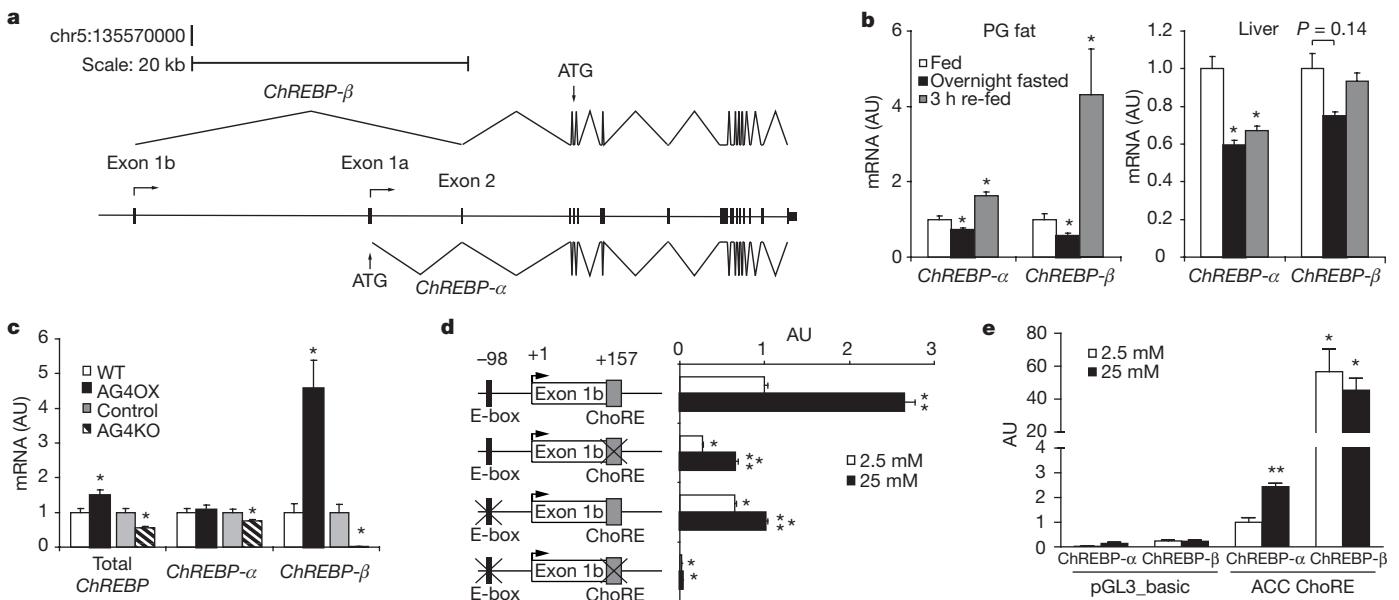


Figure 4 | Expression of the novel *ChREBP*- β isoform is regulated in a glucose- and *ChREBP*-dependent manner. **a**, Model of *ChREBP*- α and *ChREBP*- β gene structure with indication of splice sites and translational start sites (ATG). **b**, Regulation of *ChREBP*- α and *ChREBP*- β mRNA expression in perigonadal (PG) fat and liver of 10-week-old, female mice with fasting and re-feeding ($n = 6$ per group). * $P < 0.05$ compared to fed group. **c**, *ChREBP*- α and *ChREBP*- β mRNA expression in perigonadal fat from 6-week-old female AG4OX and AG4KO mice compared to littermate controls ($n = 10$ –14 per group). * $P < 0.05$ versus respective control. **d**, Glucose regulation of exon 1b

promoter-luciferase reporter and indicated *ChREBP*- β mutants, co-transfected with *ChREBP*- α and *Mlx* ($n = 3$ per group). * $P < 0.05$ compared to non-mutated exon 1b-luciferase construct in the same glucose. ** $P < 0.05$ compared to the same construct in low glucose. **e**, *ChREBP*- α and *ChREBP*- β induce an ACC1 ChoRE-luciferase reporter compared to pGL3_basic control vector in both low- and high-glucose conditions. * $P < 0.05$ compared to *ChREBP*- α in the same glucose levels, ** $P < 0.05$ compared to *ChREBP*- α in low-glucose conditions. Values are means \pm standard error.

alternative first exon, exon 1b (Fig. 4a and Supplementary Figs 16–18), and cloned an mRNA species transcribed from this alternative promoter in which exon 1b is spliced to exon 2, bypassing exon 1a and retaining the remainder of exons present in the canonical *ChREBP* mRNA species (Fig. 4a).

Physiological regulation of ChREBP isoforms

To determine whether the two distinct *ChREBP* mRNA species are physiologically regulated, we performed quantitative polymerase chain reaction (qPCR) with primers specific for exon 1a (*ChREBP-α*) or exon 1b (*ChREBP-β*). In WAT, *ChREBP-α* expression declines modestly with overnight fasting, and increases 64% after 3 h of re-feeding a chow diet compared to mice fed ad libitum (Fig. 4b). In contrast, adipose *ChREBP-β* declines 45% after an overnight fast and increases 4.3-fold with re-feeding compared to ad libitum feeding. Thus, *ChREBP-β* expression is more highly regulated in WAT than *ChREBP-α* with fasting and re-feeding. In liver, *ChREBP-α* expression declines 40% and *ChREBP-β* expression tends to decline (25%) with overnight fasting. After 3 h of re-feeding, *ChREBP-α* remains suppressed whereas *ChREBP-β* returns to baseline. After 6 h of re-feeding, liver *ChREBP-α* levels return to normal (data not shown). Thus, in liver, the time course of *ChREBP-α* and *ChREBP-β* regulation differ.

To determine whether alterations in glucose flux selectively regulate *ChREBP-α* or *ChREBP-β*, we examined the expression of the two isoforms in AG4OX and AG4KO WAT. Total *ChREBP* expression increases 50% in AG4OX and decreases 50% in AG4KO WAT (Fig. 4d). Adipose *ChREBP-α* expression remains unchanged in AG4OX and declines modestly in AG4KO. In contrast, in WAT, *ChREBP-β* increases 4.6-fold in AG4OX mice, and decreases 97% in AG4KO (Fig. 4c). Considering the difference in magnitude of change of total *ChREBP* compared to *ChREBP-β*, *ChREBP-β* is probably much less abundant than *ChREBP-α*. Although expression of both *ChREBP-α* and *ChREBP-β* are nutritionally regulated (Fig. 4b), *ChREBP-β* alone responds robustly to GLUT4-mediated changes in glucose flux (Fig. 4c).

Glucose-activated ChREBP-α induces ChREBP-β

We investigated the molecular mechanism by which *ChREBP-β* expression is regulated in response to changes in adipose GLUT4 expression. Expression of a *ChREBP-β*-promoter-luciferase construct is markedly increased in a glucose-dependent manner with co-transfection of *ChREBP-α* and MLX (Fig. 4d). MLX, the dimerization partner for ChREBP, is required for ChREBP transcriptional activity²⁹. Co-transfection of both *ChREBP-α* and MLX was required for induction of *ChREBP-β*-promoter-luciferase activity in high glucose conditions (Supplementary Fig. 19), indicating that *ChREBP-β* requires transactivation by *ChREBP-α* for expression. Basal and glucose-stimulated expression is attenuated with deletion of either the ChoRE or the upstream E-box and abolished entirely with deletion of both (Fig. 4d). Glucose and *ChREBP-α*/MLX had no effect on the expression of a luciferase reporter containing 5 kb of the *ChREBP-α* promoter, indicating that *ChREBP-α* does not regulate its own expression (data not shown). Thus, *ChREBP-α* specifically regulates expression of *ChREBP-β*.

The translational start site for *ChREBP-α* resides in exon 1a and translation from this site produces an 864-amino-acid protein. No translational start site is present in exon 1b and translation beginning at the next start site, in exon 4, produces a 687-amino-acid protein in which two nuclear export signals, a nuclear localization signal, and a domain that inhibits ChREBP transcriptional activity in low-glucose conditions are deleted (Supplementary Fig. 20)^{31,35}. N-terminally deleted ChREBP mutants demonstrate increased nuclear localization and enhanced transcriptional activity in both low- and high-glucose conditions³¹. Using an ACC1-ChoRE-luciferase reporter construct³⁶, we compared the transcriptional activity of *ChREBP-α* and *ChREBP-β* (Fig. 4e). High glucose increased *ChREBP-α* transcriptional activity

2.4-fold over low glucose. *ChREBP-β* activity in either low or high glucose is increased ~20-fold compared to *ChREBP-α* activity in high glucose, and shows no glucose regulation. The elevated *ChREBP-β* transcriptional activity may result in part from absence of the N-terminal inhibitory domain or its constitutive nuclear localization in low- and high-glucose conditions (Supplementary Fig. 21). These results suggest a feed-forward mechanism in which glucose-stimulated *ChREBP-α* transactivates expression of *ChREBP-β* (Fig. 4d and Supplementary Fig. 21), a more potent activator of ChREBP transcriptional targets (Fig. 4e).

Pathophysiological regulation of ChREBP-α/β

In mice fed a HFD, *ChREBP-β* expression declines in WAT whereas *ChREBP-α* expression remains unchanged (Fig. 5a), consistent with the possibility that downregulation of adipose *ChREBP-β* contributes to insulin resistance.

In human WAT, expression of both *ChREBP-α* and *ChREBP-β* correlate with insulin sensitivity (Fig. 5b). Multiple linear regression analysis demonstrates that expression of *ChREBP-β* ($\beta = 0.454$, $P = 0.016$) and not *ChREBP-α* ($\beta = 0.136$, $P = 0.356$) predicts insulin sensitivity (Supplementary Table 5).

Implications

Adipocytes are key regulators of whole-body energy homeostasis and metabolic function^{2,4}. Our prior work demonstrating that adipose GLUT4 expression regulates systemic insulin sensitivity^{4,5} indicated that adipocytes are capable of sensing and coordinating responses to changes in glucose availability. The elucidation here of the molecular mechanism by which glucose-mediated ChREBP activation transactivates expression of a novel, potent *ChREBP-β* (Supplementary Fig. 22) isoform reveals the complexity underlying cellular responses to nutrient availability. Our data demonstrate that adipose tissue ChREBP has a key role in integrating adipocyte and whole-body metabolic function and this may be mediated by transcriptional regulation of the potent *ChREBP-β* isoform.

The beneficial effects of adipose tissue ChREBP on glucose homeostasis may result from upregulation of adipose tissue DNL and/or other ChREBP-dependent actions such as changing the metabolic fate of glucose after it enters adipocytes and/or regulating a novel adipokine. The strong, direct correlation between adipose *ChREBP* expression and insulin sensitivity in humans suggests that adipose ChREBP may be involved in regulating whole-body insulin action. The beneficial metabolic effects of increased DNL in adipose tissue contrast with the adverse effects in the liver. DNL in liver increases intrahepatic triglyceride content and may contribute to insulin resistance and the metabolic syndrome^{37,38}. In summary, our data support the importance of adipose ChREBP in regulating insulin

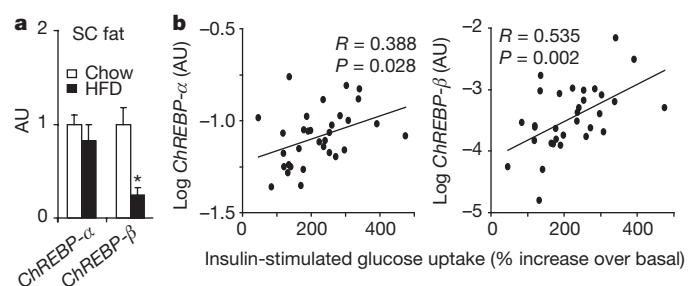


Figure 5 | Adipose tissue *ChREBP-β* expression predicts insulin sensitivity. **a**, *ChREBP-α* and *ChREBP-β* mRNA expression in subcutaneous (SC) fat of 4-month-old male mice on chow or HFD ($n = 10$ –14 per group). * $P < 0.05$ compared to chow-fed. **b**, mRNA expression of *ChREBP-β* in subcutaneous fat correlates more highly with insulin sensitivity (percentage increase in insulin-stimulated glucose uptake over basal measured during a euglycaemic and hyperinsulinaemic clamp procedure) than *ChREBP-α* in obese, non-diabetic, BMI-matched humans ($n = 38$). Values are means \pm standard error.

action and glucose homeostasis in human obesity and diabetes. These data suggest that selective activation of adipose tissue ChREBP, and specifically ChREBP- β , could be an effective therapeutic strategy for preventing and treating type 2 diabetes and obesity-related metabolic diseases.

METHODS SUMMARY

AG4OX, AG4KO and ChREBP KO mice were described previously^{4–6} except that AG4KO mice were generated using adiponectin-Cre-expressing mice³⁹ rather than aP2-Cre-expressing mice for all studies except for DNA microarrays. Phenotypic analyses were performed as described^{4,5,40}. All mouse studies were conducted in accordance with federal guidelines and were approved by the BIDMC Institutional Animal Care and Use Committee. For microarray analysis, RNA was isolated from perigonadal adipose tissue and the cDNA was analysed on the Affymetrix MG-U74-A.v2 Genechip. Quantitative real-time PCR was performed on the Applied Biosystems 7900 HT using SYBR Green PCR Master Mix. Cloning, mutagenesis, and luciferase assays were performed by standard methods. Human subjects were recruited as described^{37,41}. Human studies were approved by the Human Studies Committee of Washington University School of Medicine or the Ethics Committee of the University of Leipzig.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 December 2010; accepted 23 February 2012.

Published online 1 April 2012.

- Shepherd, P. R. & Kahn, B. B. Glucose transporters and insulin action—implications for insulin resistance and diabetes mellitus. *N. Engl. J. Med.* **341**, 248–257 (1999).
- Attie, A. D. & Scherer, P. E. Adipocyte metabolism and obesity. *J. Lipid Res.* **50** (Suppl.), S395–S399 (2009).
- Boden, G. Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* **46**, 3–10 (1997).
- Abel, E. D. *et al.* Adipose-selective targeting of the *GLUT4* gene impairs insulin action in muscle and liver. *Nature* **409**, 729–733 (2001).
- Shepherd, P. R. *et al.* Adipose cell hyperplasia and enhanced glucose disposal in transgenic mice overexpressing *GLUT4* selectively in adipose tissue. *J. Biol. Chem.* **268**, 22243–22246 (1993).
- Iizuka, K., Bruick, R. K., Liang, G., Horton, J. D. & Uyeda, K. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proc. Natl Acad. Sci. USA* **101**, 7281–7286 (2004).
- Roberts, R. *et al.* Markers of *de novo* lipogenesis in adipose tissue: associations with small adipocytes and insulin sensitivity in humans. *Diabetologia* **52**, 882–890 (2009).
- Hoffstedt, J., Forster, D. & Lofgren, P. Impaired subcutaneous adipocyte lipogenesis is associated with systemic insulin resistance and increased apolipoprotein B/AI ratio in men and women. *J. Intern. Med.* **262**, 131–139 (2007).
- Kursawe, R. *et al.* Cellularity and adipogenic profile of the abdominal subcutaneous adipose tissue from obese adolescents: association with insulin resistance and hepatic steatosis. *Diabetes* **59**, 2288–2296 (2010).
- Ranganathan, G. *et al.* The lipogenic enzymes DGAT1, FAS, and LPL in adipose tissue: effects of obesity, insulin resistance, and TZD treatment. *J. Lipid Res.* **47**, 2444–2450 (2006).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* **109**, 1125–1131 (2002).
- Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
- Repa, J. J. *et al.* Regulation of mouse sterol regulatory element-binding protein-1c gene (*SREBP-1c*) by oxysterol receptors, LXR α and LXR β . *Genes Dev.* **14**, 2819–2830 (2000).
- Cha, J. Y. & Repa, J. J. The liver X receptor (LXR) and hepatic lipogenesis. The carbohydrate-response element-binding protein is a target gene of LXR. *J. Biol. Chem.* **282**, 743–751 (2007).
- Denechaud, P. D., Girard, J. & Postic, C. Carbohydrate responsive element binding protein and lipid homeostasis. *Curr. Opin. Lipidol.* **19**, 301–306 (2008).
- Pashkov, V. *et al.* Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner. *J. Biol. Chem.* **286**, 15116–15125 (2011).
- Minn, A. H., Hafele, C. & Shalev, A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces β -cell apoptosis. *Endocrinology* **146**, 2397–2405 (2005).
- Ma, L., Robinson, L. N. & Towle, H. C. ChREBP-Mlx is the principal mediator of glucose-induced gene expression in the liver. *J. Biol. Chem.* **281**, 28721–28730 (2006).
- Kuriyama, H. *et al.* Compensatory increase in fatty acid synthesis in adipose tissue of mice with conditional deficiency of SCAP in liver. *Cell Metab.* **1**, 41–51 (2005).
- Cao, H. *et al.* Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* **134**, 933–944 (2008).
- Carvalho, E., Kotani, K., Peroni, O. D. & Kahn, B. B. Adipose-specific overexpression of *GLUT4* reverses insulin resistance and diabetes in mice lacking *GLUT4* selectively in muscle. *Am. J. Physiol. Endocrinol. Metab.* **289**, E551–E561 (2005).
- Ahima, R. S. & Flier, J. S. Leptin. *Annu. Rev. Physiol.* **62**, 413–437 (2000).
- Nadler, S. T. *et al.* The expression of adipogenic genes is decreased in obesity and diabetes mellitus. *Proc. Natl Acad. Sci. USA* **97**, 11371–11376 (2000).
- Caesar, R. *et al.* A combined transcriptomics and lipidomics analysis of subcutaneous, epididymal and mesenteric adipose tissue reveals marked functional differences. *PLoS ONE* **5**, e11525 (2010).
- Gnudi, L., Tozzo, E., Shepherd, P. R., Bliss, J. L. & Kahn, B. B. High level overexpression of glucose transporter-4 driven by an adipose-specific promoter is maintained in transgenic mice on a high fat diet, but does not prevent impaired glucose tolerance. *Endocrinology* **136**, 995–1002 (1995).
- Sanchez, J., Palou, A. & Pico, C. Response to carbohydrate and fat refeeding in the expression of genes involved in nutrient partitioning and metabolism: striking effects on fibroblast growth factor-21 induction. *Endocrinology* **150**, 5341–5350 (2009).
- Shimano, H. *et al.* Elevated levels of SREBP-2 and cholesterol synthesis in livers of mice homozygous for a targeted disruption of the SREBP-1 gene. *J. Clin. Invest.* **100**, 2115–2124 (1997).
- Stoeckman, A. K., Ma, L. & Towle, H. C. Mlx is the functional heteromeric partner of the carbohydrate response element-binding protein in glucose regulation of lipogenic enzyme genes. *J. Biol. Chem.* **279**, 15662–15669 (2004).
- Kabashima, T., Kawaguchi, T., Wadzinski, B. E. & Uyeda, K. Xylulose 5-phosphate mediates glucose-induced lipogenesis by xylulose 5-phosphate-activated protein phosphatase in rat liver. *Proc. Natl Acad. Sci. USA* **100**, 5107–5112 (2003).
- Li, M. V., Chang, B., Imamura, M., Pongvarin, N. & Chan, L. Glucose-dependent transcriptional regulation by an evolutionarily conserved glucose-sensing module. *Diabetes* **55**, 1179–1189 (2006).
- Iizuka, K., Takeda, J. & Horikawa, Y. Hepatic overexpression of dominant negative Mlx improves metabolic profile in diabetes-prone C57BL/6J mice. *Biochem. Biophys. Res. Commun.* **379**, 499–504 (2009).
- Shih, H. M., Liu, Z. & Towle, H. C. Two CACGTG motifs with proper spacing dictate the carbohydrate regulation of hepatic gene transcription. *J. Biol. Chem.* **270**, 21991–21997 (1995).
- Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl Acad. Sci. USA* **99**, 8695–8700 (2002).
- Fukasawa, M., Ge, Q., Wynn, R. M., Ishii, S. & Uyeda, K. Coordinate regulation/localization of the carbohydrate responsive binding protein (ChREBP) by two nuclear export signal sites: discovery of a new leucine-rich nuclear export signal site. *Biochem. Biophys. Res. Commun.* **391**, 1166–1169 (2010).
- Tsatsos, N. G. & Towle, H. C. Glucose activation of ChREBP in hepatocytes occurs via a two-step mechanism. *Biochem. Biophys. Res. Commun.* **340**, 449–456 (2006).
- Fabbri, E. *et al.* Intrahepatic fat, not visceral fat, is linked with metabolic complications of obesity. *Proc. Natl Acad. Sci. USA* **106**, 15430–15435 (2009).
- Donnelly, K. L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J. Clin. Invest.* **115**, 1343–1351 (2005).
- Eguchi, J. *et al.* Transcriptional control of adipose lipid handling by IRF4. *Cell Metab.* **13**, 249–259 (2011).
- Kotani, K., Peroni, O. D., Minokoshi, Y., Boss, O. & Kahn, B. B. *GLUT4* glucose transporter deficiency increases hepatic lipid production and peripheral lipid utilization. *J. Clin. Invest.* **114**, 1666–1675 (2004).
- Klötting, N. *et al.* Serum retinol-binding protein is more highly expressed in visceral than in subcutaneous adipose tissue and is a marker of intra-abdominal fat mass. *Cell Metab.* **6**, 79–87 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Towle for providing reagents and discussions. We thank P. Pissios, I. Astapova, E. Rosen and the Rosen laboratory for discussions, K. Uyeda for providing ChREBP KO mice, and E. Shu for technical assistance. This work was supported by NIH R37 DK43051 (B.B.K.), K08 DK076726 (M.A.H.), BADERC DK057521 (B.B.K., M.A.H.), BNORC DK046200 (M.A.H.), the Picower and JPB Foundations (B.B.K.), a Fellowship from the Radcliffe Institute for Advanced Study (B.B.K.), DK056341 (Nutrition and Obesity Research Unit) (S.K.), DK037948 (S.K.), and the Deutsche Forschungsgemeinschaft DFG, KFO152, BL833/1 (M.B.).

Author Contributions M.A.H. and B.B.K. planned the experiments, interpreted the data, and wrote the paper. M.A.H. designed and performed the molecular biology experiments to discover ChREBP- β and executed all other experiments in the paper except the human studies. O.D.P. and J.V. assisted with experimental design and provided technical support. N.A.A., M.R.S., M.B. and S.K. performed the human studies.

Author Information The ChREBP- β sequence data has been deposited in NCBI Genbank under accession number JQ437838. The microarray data has been deposited in NCBI Gene Expression Omnibus under accession number GSE35378. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.B.K. (bkahn@bidmc.harvard.edu).

METHODS

Animal studies. Generation and initial metabolic characterization of the AG4OX, AG4KO and ChREBP KO mice were previously described^{4–6} except that AG4KO mice were generated using adiponectin-Cre-expressing mice³⁹ rather than aP2-Cre-expressing mice for all studies except for the DNA microarrays. Wild type and adiponectin-Cre littermates were used as controls for AG4KO mice. Mice were housed at Beth Israel Deaconess Medical Center with a 14/10 light–dark cycle and were fed standard chow (Formulab 5008) or HFD (Harlan-Teklad TD.93075) ad libitum. All studies were performed on age- and sex-matched littermates. Phenotypic analyses were performed as described^{4,5,40}. Blood collections were performed by submandibular vein or tail vein bleeding. Body composition was measured by NMR (Echo Medical Systems). Glucose tolerance tests were performed by injection of 1 mg glucose per kg body weight intraperitoneally after 5 h food removal. For the fasting glycaemic time course experiment, food was removed at 8:00. Mice were killed by CO₂ euthanasia or decapitation and serum was collected, and tissues were harvested, snap frozen in liquid nitrogen, and stored at –80 °C for processing. Mouse studies were conducted in accordance with federal guidelines and were approved by the BIDMC Institutional Animal Care and Use Committee.

Human studies. For the cross-sectional cohort, adipose tissue samples were obtained from 123 men ($n = 64$) and women ($n = 59$) who underwent open abdominal surgery for gastric banding, cholecystectomy, appendectomy, weight reduction surgery, abdominal injuries or explorative laparotomy. These patients represent a normal glucose tolerant subset of patients previously described⁴¹. All subjects had a stable weight, defined as the absence of fluctuations of >2% of body weight for at least 3 months before surgery. Patients with malignant diseases or any acute or chronic inflammatory disease, as determined by a leukocyte count of >7,000 Gpt l^{–1}, C-reactive protein levels of >50 mg l^{–1} or clinical signs of infection, were excluded from the study. After resection, samples of visceral and subcutaneous adipose tissue were immediately frozen in liquid nitrogen. Euglycaemic and hyperinsulinaemic clamp studies were performed before surgery in subjects undergoing bariatric surgery or 3 months after surgery in subjects who underwent non-elective surgeries. The clamp study was performed with an insulin infusion rate of 20 mU kg^{–1} min^{–1}, as described, and the glucose disposal rate was defined as the glucose infusion rate during the last 30 min of the study⁴¹. The study was approved by the ethics committee of the University of Leipzig. All participants gave written informed consent before taking part in the study.

For the obese cohort, 38 obese subjects (10 men, 28 women; 41 ± 11 years old) who had normal oral glucose tolerance were studied. Subjects were sedentary (that is, participated in regular exercise <1 h per week and ≤ 1 time per week) and weight stable. All subjects provided written, informed consent before participating in the study, which was approved by the Human Studies Committee of Washington University School of Medicine in St Louis.

Subjects were admitted to the Clinical Research Unit at Washington University School of Medicine in the evening before the clamp procedure. The next morning, after subjects fasted for 12 h overnight, a hyperinsulinaemic and euglycaemic clamp procedure in conjunction with [6,6-²H₂]glucose tracer infusion was performed, as described previously³⁷. Euglycaemia was maintained at a blood glucose concentration of approximately 5.6 mmol l^{–1} (100 mg dl^{–1}) by infusing 20% dextrose enriched to 2.5% with [6,6-²H₂]glucose. Skeletal muscle insulin sensitivity was determined as the relative increase in the rate of glucose uptake during insulin infusion (50 mU m^{–2} min^{–1}).

Microarray analysis. Total RNA from epididymal adipose tissue was extracted using the RNeasy Mini Kit from Qiagen from three mice from each of four genotypes: aP2-Cre transgenic littermates (controls for AG4KO mice), AG4KO mice; FVB littermates (controls for AG4OX) and AG4OX. RNA from each mouse was hybridized on an Affymetrix MG-U74-A.v2 Genechip microarray. Affymetrix gene chip hybridization and analysis were performed at the Genomics Core Facility of the Beth Israel Deaconess Medical Center. Genome-wide expression analysis of the microarray data were performed using gene set enrichment analysis (GSEA)¹¹ and the s2.mgu74av2.gmt gene set database available through the Broad Institute.

Gene expression analysis. Mouse tissues were harvested following CO₂ euthanasia or decapitation, snap frozen in liquid nitrogen, and stored at –80 °C for processing. Total RNA was extracted from frozen tissue with TRI Reagent (Molecular Research Center). Reverse transcription was performed using the Advantage RT-for-PCR kit (Clontech). Quantitative PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) in a 7900 HT thermocycler (Applied Biosystems). SDS 2.3 (Applied Biosystems) was used for calculation of cycle thresholds. Relative expression levels were determined using the standard curve method with normalization of target gene expression levels to

18s ribosomal RNA (18s) or TATA box binding protein (*Tbp*). See Supplementary Table 6 for primer sequences.

For the cross-sectional cohort, total RNA was isolated from subcutaneous adipose tissue samples using TRIzol (Life Technologies) and cDNA was synthesized with standard reagents (Life Technologies). Quantitative PCR was performed using Brilliant SYBR Green QPCR Core Reagent Kit (Stratagene) in a PRISM 7900 thermocycler (Applied Biosystems). Relative expression levels were determined using the standard curve method with normalization of target gene expression levels to 18 s.

For the obese cohort, total RNA was isolated from subcutaneous adipose tissue samples using TRIzol (Invitrogen) and cDNA was synthesized using Taqman Reverse Transcription (Applied Biosystems). Quantitative PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) on the ABI 7500 Real-Time PCR System (Applied Biosystems). Relative expression levels were determined using the 2^{–ΔCt} method with normalization of target gene expression levels to ribosomal protein, large, P0 (36B4) gene.

Western blotting. For adipose tissue lysates, aliquots of frozen adipose tissue were homogenized on ice in RIPA buffer supplemented with phospho-preserving and anti-protease agents (sodium fluoride, sodium pyrophosphate, sodium orthovanadate, PMSF, aprotinin and leupeptin). For adipose tissue nuclear lysates, an adipose tissue sample was ground under liquid nitrogen, then dounce homogenized in low salt buffer (potassium chloride 10 mM; HEPES pH 7.9, 10 mM; EDTA 0.1 mM; EGTA 0.1 mM; DTT 1 mM) with protease inhibitors. Homogenates were centrifuged at 1,000g for 10 min at 4 °C and nuclear pellets were collected and washed with low salt buffer. The nuclear pellets were resuspended in RIPA buffer and vortexed vigorously. Nuclear lysates from 3–4 mice from the same genotype were pooled for further processing. Protein concentrations were assayed using the BCA assay. Equal amounts of protein were loaded and transferred to nitrocellulose membranes. Membranes were probed with antibodies against acetyl-CoA carboxylase (Cell Signaling), fatty acid synthase (Cell Signaling), GLUT4 (provided by H. Haspel), SREBP1 (Novus Biologicals), and lamin B (Santa Cruz) and PI3K p85 (Upstate). Film images were scanned (Epson Expression 10000 XL) and results were quantified with ImageQuant TL software (GE).

In vivo fatty acid synthesis. Conscious, freely moving mice in the fed state were injected intraperitoneally with 5 mCi of ³H₂O (MP Biomedicals) and 10 μCi of [U-¹⁴C]-glucose (Perkin Elmer) and killed 1 h later using an overdose of ketamine and xylazine (160 and 24 mg kg^{–1} intraperitoneally, respectively). Samples of plasma were obtained serially at 5, 10, 30 and 60 min after injection for measurement of the plasma glucose levels and the specific activity of plasma water and glucose. At sacrifice, tissues were rapidly removed, frozen in liquid N₂ and stored at –80 °C for processing. Lipids were extracted by the Folch method with chloroform/methanol (2:1). Fatty acids were obtained by saponification and extraction with petroleum ether. Incorporation of ³H and ¹⁴C into saponified fatty acids were measured. The rate of synthesis was calculated as a molar rate based on the estimate that 13.3 mol of H₂O are incorporated into each newly synthesized C16 fatty acid⁴². The rate of glucose incorporation into newly synthesized fatty acids was calculated as nanomoles of ¹⁴C-glucose incorporated into fatty acids per gram tissue per minute, using ¹⁴C counts in saponified fatty acids extracted from tissue normalized by the time-averaged specific activity of glucose in plasma over the course of the hour-long experiment.

Glucose uptake in isolated adipocytes. Adipocytes were isolated from perigonadal fat pads and glucose uptake was measured as previously described⁵. Briefly, perigonadal fat pads were digested with collagenase (1 mg ml^{–1}) and cells were incubated at 37 °C with constant shaking in Krebs-Ringer-HEPES (30 mM) buffer (pH 7.4) with 2% bovine serum albumin, 200 nM adenosine, and without (basal) or with (insulin-stimulated) 80 nM insulin. After a 30-min incubation with or without insulin, U-¹⁴C glucose (3 μM) was added for 60 min and the reaction was terminated by separating cells from media by spinning the suspension through dinonyl-phthalate oil. A portion of isolated adipocytes from each sample were fixed with osmic acid and counted in a Coulter counter to normalize glucose uptake per cell.

Identification of ChoRE and ChREBP exon 1b. A position-weighted matrix representative of a consensus ChoRE was generated from sequences for 16 mouse, human, or rat ChoREs previously identified and experimentally validated (Supplementary Fig. 15). Genomic sequence 20 kb upstream and downstream of the mouse *ChREBP* transcriptional start site was obtained from the UCSC Genome Browser (Mouse July 2007 (NCBI37/mm9) Assembly)^{43,44} and was scanned using the Transcriptional Element Search System (<http://www.cbil.upenn.edu/tess>) and the positional weight described above. Genomic sequence in the exon 1b region was obtained through the UCSC genome browser and comparisons across species were performed with Megalign (Dnastar) using the Clustal W algorithm. Analysis of histone methylation marks was performed using

the Ensembl browser and database^{45,46}. *ChREBP* exon 1b was cloned from RNA prepared from AG4OX adipose tissue by 5' rapid amplification of cDNA ends (GeneRacer, Invitrogen). A full-length *ChREBP-β* mRNA species was cloned using GeneRacer reagents. All cloned products were verified by sequencing.

Functional analysis of ChoREs. PCR of bacterial artificial chromosome containing the mouse *ChREBP* gene (CH29-535J6, Children's Hospital Oakland Research Institute) was used to generate the mouse *ChREBP-β* promoter sequence (−631 to +224, in reference to exon 1b transcriptional start site), which was cloned into the pGL3_{basic} reporter vector. The putative ChoRE sequence (+157 to +174) and E-box (−98 to −92) were deleted by site-directed mutagenesis (QuickChange II Site-Directed Mutagenesis Kit, Agilent). The exon 1b promoter_pGL3, mutant pGL3 constructs, or an empty pGL3_{basic} control were co-transfected using Lipofectamine LTX (Invitrogen) in HEK293T cells with Flag-tagged ChREBP-α (ζ-isoform) and HA-tagged MLX, both in the CMV4 vector²⁹, along with a renilla luciferase control. Following transfection, cells were cultured in DMEM with high (25 mM) glucose. Twenty-four hours after transfection, cells were cultured in DMEM containing low (2.5 mM) versus high (25 mM) glucose. Cells were collected and firefly luciferase activity was measured 24 h after the media change and normalized to renilla luciferase activity (Dual Luciferase Assay System, Promega). All experiments were repeated at least three times with comparable results.

Functional analysis of ChREBP-β. Flag-tagged ChREBP-β was generated by deletion of N-terminal sequence in the mouse Flag-tagged ChREBP-α (GenBank accession AF245475) in the CMV4 vector²⁹ by site-directed mutagenesis (QuickChange II Site-Directed Mutagenesis Kit, Agilent). ChREBP-β sequence was confirmed by sequencing. Flag-tagged ChREBP-α, Flag-tagged ChREBP-β, or empty pGL3_{basic} control were co-transfected with HA-tagged MLX and a reporter plasmid containing a promoter region consisting of two copies of the ACC1 ChoRE fused to the firefly luciferase gene³⁶. Following transfection, cells were cultured in DMEM with high (25 mM) glucose. Twenty-four hours after

transfection, cells were cultured in DMEM containing low (2.5 mM) or high (25 mM) glucose. Cells were collected and firefly luciferase activity was measured 24 h after the media change and normalized to renilla luciferase activity (Dual Luciferase Assay System, Promega). Similar results were also obtained when normalizing luciferase activity to ChREBP-α or ChREBP-β protein levels, respectively (data not shown). All experiments were repeated at least three times with comparable results.

Statistics. All values are given as means ± standard error. Differences between two groups were assessed using unpaired two-tailed Student's *t*-tests unless otherwise specified in figure legends. Comparisons between multiple groups were performed by ANOVA with pairwise comparisons and adjustment for multiple comparisons by the Tukey test unless otherwise specified in figure legends. For analyses with non-normal distributions or unequal variances, ANOVA and pairwise comparisons were performed on log-transformed values. All data presented in the figures are of non-transformed values. Owing to the large variance and non-normal distribution of *GLUT4* expression in the obese human cohort (Fig. 3i), log mRNA expression levels were analysed and presented. Statistical analyses including linear regressions were performed with SPSS 8.0.

42. Hems, D. A., Rath, E. A. & Verrinder, T. R. Fatty acid synthesis in liver and adipose tissue of normal and genetically obese (ob/ob) mice during the 24-hour cycle. *Biochem. J.* **150**, 167–173 (1975).
43. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
44. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2010).
45. Rosenbloom, K. R. *et al.* ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res.* **38**, D620–D625 (2010).
46. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).

A unique regulatory phase of DNA methylation in the early mammalian embryo

Zachary D. Smith^{1,2,3*}, Michelle M. Chan^{1,4*}, Tarjei S. Mikkelsen^{1,2}, Hongcang Gu¹, Andreas Gnirke¹, Aviv Regev^{1,5} & Alexander Meissner^{1,2,3}

DNA methylation is highly dynamic during mammalian embryogenesis. It is broadly accepted that the paternal genome is actively depleted of 5-methylcytosine at fertilization, followed by passive loss that reaches a minimum at the blastocyst stage. However, this model is based on limited data, and so far no base-resolution maps exist to support and refine it. Here we generate genome-scale DNA methylation maps in mouse gametes and from the zygote through post-implantation. We find that the oocyte already exhibits global hypomethylation, particularly at specific families of long interspersed element 1 and long terminal repeat retroelements, which are disparately methylated between gametes and have lower methylation values in the zygote than in sperm. Surprisingly, the oocyte contributes a unique set of differentially methylated regions (DMRs)—including many CpG island promoters—that are maintained in the early embryo but are lost upon specification and absent from somatic cells. In contrast, sperm-contributed DMRs are largely intergenic and become hypermethylated after the blastocyst stage. Our data provide a genome-scale, base-resolution timeline of DNA methylation in the pre-specified embryo, when this epigenetic modification is most dynamic, before returning to the canonical somatic pattern.

Cytosine methylation in mammals is an epigenetic modification that is largely restricted to CpG dinucleotides and serves multiple critical functions, including stable repression of target promoters, maintaining genomic integrity, establishing parent-specific imprinting patterns, and silencing endogenous retrotransposon activity^{1,2}. In somatic tissues, CpG methylation exhibits global patterns based on relative CpG density: CpG islands at housekeeping or developmental promoters are largely unmethylated, whereas non-regulatory CpGs distributed elsewhere in the genome are largely methylated^{1,3}. This landscape is relatively static across all somatic tissues, where most of the methylated CpGs are pre-established and inherited through cell division. Generally, only a small fraction of CpGs switch their methylation status as part of an orchestrated regulatory event^{4–7}.

DNA methylation is much more dynamic during mouse germ cell and pre-implantation development. The classical model postulates that at fertilization, a targeted, although widespread, catalytic process actively removes DNA methylation contributed by the paternal gamete. Recent evidence implicates a demethylation mechanism that transitions through a hydroxymethylated intermediate that is catalysed by the Tet3 member of the Tet family^{8,9}. However, only a portion of hydroxylated targets seems to be actively catalysed to complete demethylation, and the identity of these targets remains unknown¹⁰. After fertilization, there appears to be a passive loss of global DNA methylation levels that continues until the blastocyst stage, where the inner cell mass (ICM) that gives rise to the embryo proper is first specified (reviewed by ref. 11). Recent evidence indicates that this passive depletion may also be facilitated in part by Tet-enzyme-mediated hydroxylation¹⁰. After specification of the ICM, the embryo implants into the uterine lining in concert with gastrulation, which

co-occurs with global remethylation of the genome that is believed to contribute to lineage restriction and the loss of cellular potency^{12,13}.

Unfortunately, on a quantitative, genome-wide scale, little is known about the specific dynamics of cytosine methylation during these earliest developmental stages¹⁴. The classical model is drawn from observations made using either global measurements, such as immunohistochemistry, or from limited analysis at individual loci^{11,12,15–22}. Key questions about DNA methylation patterns in early development remain open, including which genomic features are specifically targeted, as well as the identities of DMRs inherited from either gamete beyond known imprint control regions (ICRs). Here we use genomic high-resolution methylation profiling^{23,24} to gain insight into the underlying mechanisms and regulatory principles of DNA methylation as it functions in early mammalian development.

Genome-scale methylation maps of murine embryogenesis

To generate a global and high-resolution view of early mammalian DNA methylation dynamics, we collected oocytes and sperm, as well as zygote, 2-, 4- and 8-cell cleavage stage embryos, the early ICM and embryonic day (E)6.5/7.5 post-implantation embryos (Fig. 1a and Supplementary Figs 1 and 2). All samples were extensively washed and purified to remove any somatic or gametic contaminants; maternal biasing from meiotic polar bodies (representing a 1× or 0.5× static genomic contaminant, respectively) was excluded by mechanical biopsy (Supplementary Fig. 1 and Supplementary Movie 1) and was further confirmed by assessing the paternal (129X1/SvJ) to maternal (C57BL/6 × DBA/2) ratio of known single nucleotide polymorphisms (SNPs) (Supplementary Figs 3 and 4). We

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. ³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

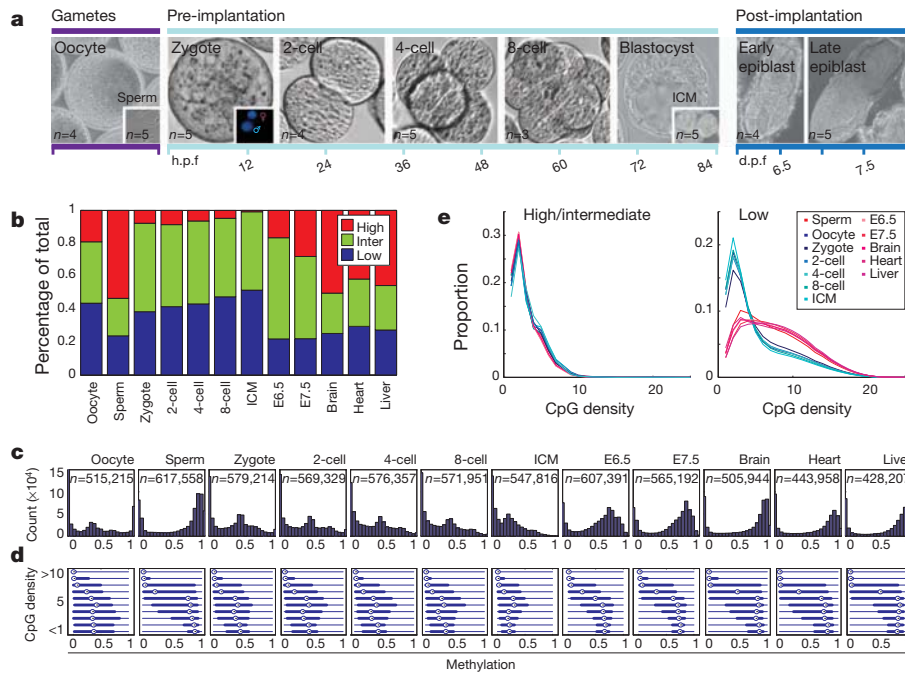


Figure 1 | Global CpG methylation dynamics across early murine embryogenesis. **a**, Samples isolated for methylation analysis with replicate number (n) highlighted. d.p.f., days post fertilization; h.p.f., hours post fertilization. **b**, Fraction of 100-bp tiles with high (≥ 0.8 , red), intermediate (inter, >0.2 and <0.8 , green) and low (≤ 0.2 , blue) methylation values. Brain, heart and liver tissue are shown for adult comparisons. **c**, Histogram of methylation values across 100-bp tiles. n is the number of tiles for each stage.

generated reduced representation bisulphite sequencing (RRBS)⁴ libraries from each stage to provide a comprehensive timeline of DNA methylation patterns during early mouse embryogenesis.

Compared to all other genome-wide profiling strategies currently available, RRBS is optimally suited for the low cell numbers that can be obtained from embryonic samples^{23–25}. Within our range of ~ 0.5 – 10 ng genomic DNA, RRBS provides the expected genomic coverage and high reproducibility (Supplementary Fig. 2). On average, we obtained the methylation status of 1,062,216 CpGs for comparative analysis (Supplementary Table 1). Bisulphite sequencing cannot distinguish between methyl- and hydroxymethylcytosine (hmC), and current methods for global profiling of hmC lack the sensitivity to investigate the pre-implantation stages in this study^{9,26–31}. Thus, we cannot draw any definitive conclusions regarding the base resolution hmC distribution, but this modification has not yet been linked to a regulatory mechanism other than to potentiate demethylation³². Given this ambiguity, regions of high mC/hmC methylation, especially those retained over multiple time points, could still be expected to function as if methylated.

The global embryonic pattern is unique

Current models postulate a phase of global hypomethylation during mammalian pre-implantation development that reaches a minimum at the morula/blastocyst stage. However, it is unknown which genomic regions are affected or how similar the embryonic methylation pattern is to the adult¹. To address these questions, we investigated the global dynamics of CpG methylation using 100-base-pair (bp) tiles (Methods). We found that oocytes are already globally hypomethylated compared to sperm (0.32 median methylation in oocyte versus 0.83 in sperm; Supplementary Fig. 5). We examined the relative proportion of genomic regions at each stage falling into high (≥ 0.8), intermediate (>0.2 and <0.8) or low (≤ 0.2) methylation categories. Notably, oocyte methylation levels more closely resemble those of early embryonic time points than the levels in sperm, post-implantation

d, Box plots of methylation values across local CpG densities highlight the difference between hypomethylated pre-implantation tissues and the somatic pattern seen in sperm, post-implantation and adult samples. Circle indicates the median, edges the 25th/75th percentile and whiskers the 2.5th/97.5th percentile. **e**, CpG density distribution for >0.2 methylation (left panel) and ≤ 0.2 methylation (right panel) tiles in stages that display somatic versus embryonic patterning (red and blue lines, respectively).

embryos, or adult tissues (Fig. 1b). We also observed a gradual increase in the fraction of tiles that exhibit intermediate and low methylation values from oocytes to the early ICM, which is consistent with loss of methylation over multiple cleavage divisions (Fig. 1b).

Sperm and post-implantation embryos show a strong inverse relationship between CpG density and methylation levels that is characteristic of somatic cells. In oocyte and pre-implantation samples, this dependence is weaker (Fig. 1c, d). In both pre- and post-implantation embryos, methylated CpGs (>0.2) tend to occur in low CpG density regions, as they do in somatic cells (Fig. 1e, left). However, the alternative relationship between higher CpG density and low methylation is not as apparent in the oocyte or the pre-implantation embryo (Fig. 1e, right). In summary, pre-implantation development represents a unique developmental period where methylation is differentially positioned and regulated before being restored and maintained in a somatic fashion.

Two transitions in early development

We next searched for substantial changes in regional DNA methylation through development. For each pair of consecutive stages, we compared methylation levels of each tile and classified it as changed if the difference exceeded 0.2 and was significant according to a false discovery rate (FDR)-corrected t -test. The most dramatic changes in DNA methylation occurred during two developmental transitions: between sperm and the zygote and between the early ICM and the post-implantation embryo (Fig. 2a). At each of these transitions, most changes were unidirectional (Fig. 2b): a gross reduction upon fertilization (mean = 0.47 decrease for 37% of tiles examined) and massive remethylation from the ICM onwards (mean = 0.46 increase in methylation at 66% of tiles). Within E6.5 and E7.5 post-implantation embryos, the methylation levels at most of the assayed tiles were stable or increased slightly (Fig. 2b). However, more subtle global changes, reflecting a gradual decrease in methylation, were observed from zygote/early cleavage through to the 8-cell stage and

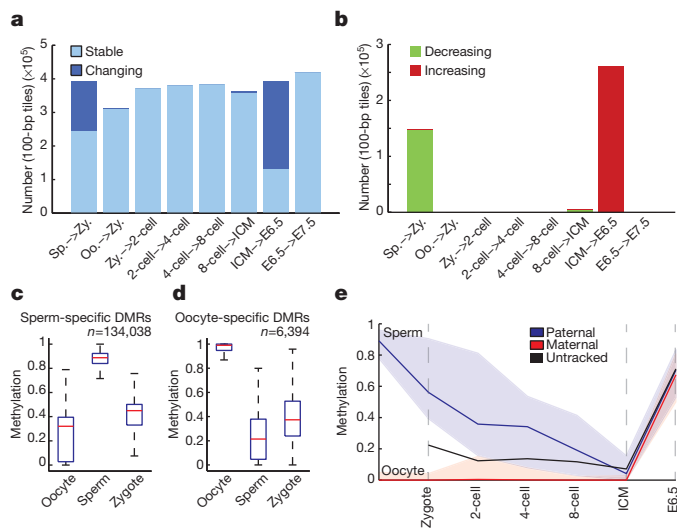


Figure 2 | Major transitions in DNA methylation levels during early development. **a**, 100-bp tiles available for pairwise comparison across consecutive embryonic stages. Tiles that remain unchanged (stable) at the indicated transitions are shown in light blue. Tiles that change by greater than 0.2 and are significant by *t*-test are highlighted in dark blue. **b**, 100-bp tiles with increasing (red) or decreasing (green) methylation levels at each consecutive transition show that major transitions are largely unidirectional. **c**, Box plot of methylation levels for sperm-specific DMRs ($n = 134,038$ tiles). Red line indicates the median, edges the 25th/75th percentile and whiskers the 2.5th/97.5th percentile. **d**, Box plot of methylation levels for oocyte-specific DMRs ($n = 6,394$ tiles) as in **c**. **e**, Seventy-four CpGs within sperm-specific DMR tiles (**c**) could be ascribed to paternal or maternal alleles and tracked across stages. Paternal CpG methylation values (blue line, median; coloured space, 25th/75th percentile) decrease by the zygote stage whereas maternal CpG methylation (red line, median; coloured space, 25th/75th percentile) remains unchanged. If untracked, these CpGs have an intermediate methylation value between those ascribed to a parent of origin (black line).

into the ICM, where methylation levels reached their lowest observed values (Fig. 1b, c).

The oocyte defines the early methylation landscape

Active demethylation is expected to occur before pronuclear fusion or the completion of DNA synthesis^{11,33}. When we compare methylation patterns between sperm and zygote, most regions in the genome show reduced methylation in the zygote with few sizeable changes in 2-cell embryos (Fig. 2b). Notably, the vast majority of tiles that are methylated at significantly different levels between gametes show higher methylation levels in sperm than in oocyte and many are reduced to levels near those of the oocyte (Fig. 2c, d). Using SNPs, we confirmed this observation by tracking 74 CpGs that fell within these tiles and could be assigned paternal- or maternal-specific values. Zygotes displayed a decrease in paternal methylation in contrast to maternally contributed CpGs, which remained unmethylated (Fig. 2e). Zygotes isolated here are likely in earlier stages of S phase, such that either a passive, replication-based mechanism could result in the synthesis of unmethylated, nascent DNA or DNA methylation could be removed by a targeted process^{10,34–36}. The similarities in methylation levels between zygote and the 2-cell stage, which represents one complete round of replication, is indicative that at least some observed demethylation is a consequence of targeted removal, but distinguishing between these two models may be complicated by the coupling of proposed base-excision repair mechanisms and DNA replication itself³⁵.

The few regions that are significantly hypermethylated in oocyte compared to sperm exhibit intermediate values in the zygote, suggesting a more direct inheritance of the allelic methylation state (Fig. 2d). The disparity in the zygotic resolution of regions that are differentially

methylated between the gametes indicates that the oocyte largely reflects the zygotic/pre-implantation methylome and prescribes its architecture (Supplementary Fig. 6). Thus, the oocyte methylome, rather than the sperm methylome, seems to be more reflective of patterns in the early embryo.

Retroelement dynamics at fertilization

Consistent with a demethylation model, we confirmed that the vast majority (96%) of tiles that are hypermethylated in sperm in our data set become less methylated in the zygote. Most of these tiles already exhibit lower methylation in the oocyte, such that additive effects could explain more subtle decreases in many regions. Interestingly, tiles exhibiting the most extreme methylation changes during the sperm to zygote transition are enriched for long interspersed elements (LINEs) ($P < 4.7 \times 10^{-184}$, $FDR < 0.05$, hypergeometric enrichment) (Fig. 3a and Supplementary Table 2). We directly estimated the methylation level for individual LINEs³⁷ surveyed by RRBS at each stage and found that changes in these elements are markedly bimodal during the sperm to zygote transition, with 18% of LINEs reducing their methylation values by over 0.45 (Fig. 3a). By comparison, 10% of captured long terminal repeat (LTR) retroelements exhibit similar levels of demethylation, but the distribution is not as clearly bimodal (Fig. 3b and Supplementary Table 3). Short interspersed elements (SINEs) are generally less methylated in sperm than other repeat classes³⁸, and also exhibit shifts in their methylation values from sperm to early embryo, but without the apparent bimodality observed for LINE elements (Supplementary Fig. 7).

Surprisingly, LINEs that changed most dramatically during the sperm to zygote transition largely consisted of two closely related families of L1 LINEs: L1Md_T and L1Md_Gf (Fig. 3c, d, $P < 4.7 \times 10^{-184}$, $P < 7.9 \times 10^{-6}$; hypergeometric enrichment test)^{37,39}. Repeats from these families had the largest and most consistent decrease, whereas those from other equally represented families, such as L1Md_A elements, showed smaller changes upon fertilization and maintained higher methylation values in both oocyte and zygote (Fig. 3e and Supplementary Fig. 8). Similarly, several LTR families showed considerable loss of methylation within the zygote (Fig. 3f, g), whereas the class II intracisternal A-particles (IAPs, Fig. 3h) did not. The latter finding is consistent with the known retention of high methylation levels of IAPs throughout cleavage^{19,22}.

Interestingly, during early development, all retrotransposons resolved identically, reaching minimal values at the ICM stage before increasing to more somatic levels by E6.5/7.5 (Fig. 3i). Thus, repeat elements exist in a less methylated state primarily in the oocyte pre-implantation stages (Supplementary Fig. 7). Bisulphite sequencing cannot address if methylated cytosines at these repeats are converted to hmCs before a subset is further targeted for complete demethylation. Some mCs may be targeted for active demethylation via this intermediate form, whereas the remaining mC/hmC residues may lose their methylation more passively through cleavage, consistent with recent metaphase immunostaining results¹⁰.

Gametes confer distinct features as DMRs

Although loss of methylation is widespread, some epigenetic information must be differentially contributed by the two gametes, including known ICRs that maintain their allele-specific methylation pattern throughout embryogenesis⁴⁰. We systematically searched for inherited DMRs contributed from either gamete by applying linear regression to all tiles that had mean methylation ≥ 0.75 in one gamete and ≤ 0.25 in the other. We identified 376 oocyte-contributed DMRs with intermediate methylation levels in the zygote ($P < 0.047$, $FDR < 0.05$, ANOVA; linear regression residual < 0.29 , $FDR < 0.1$; Fig. 4a) and 4,894 sperm-contributed DMRs (Fig. 4c). Notably, oocyte-contributed DMRs reside primarily in CpG island-containing promoters (Fig. 4b and Supplementary Table 4), whereas sperm-contributed DMRs are predominantly intergenic (Fig. 4d). The sperm- and oocyte-contributed

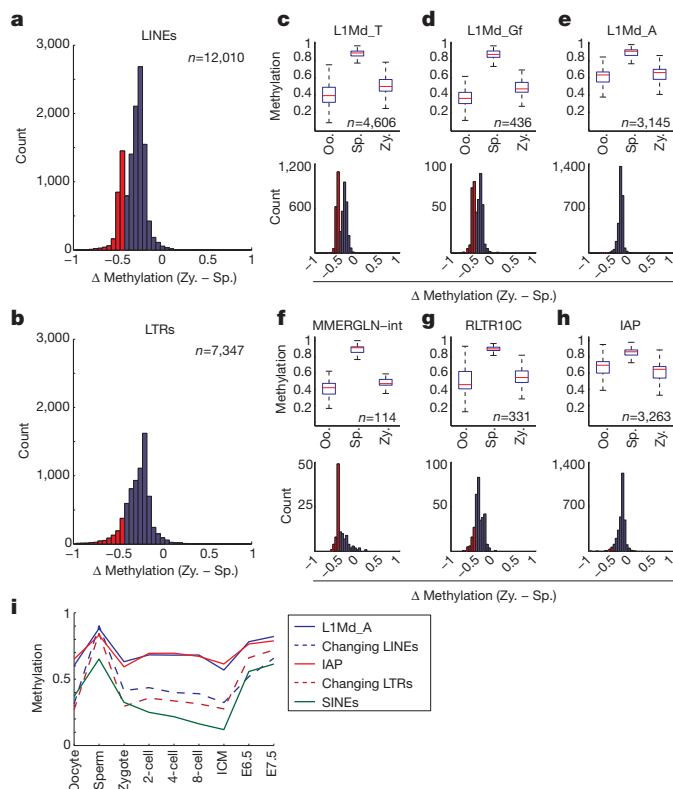


Figure 3 | Specific families of LINE and LTR retroelements exhibit the most dramatic methylation changes in the sperm to zygote transition.

a, Histogram of the difference in methylation levels, where negative values represent tiles decreasing from sperm to zygote, within LINE retroelement features that are captured by RRBS. 85% of the elements have a significant difference ($P < 0.04$, $FDR < 0.05$; t -test). The distribution is bimodal with 18% of elements displaying a change in methylation status ≥ 0.45 as highlighted in red. **b**, Differences in methylation between sperm and zygote within annotated LTR retroelements. Compared to LINEs, a smaller fraction of elements appear to be regulated by DNA demethylation (61% significant, 10% of those sampled exhibiting changes ≥ 0.45 as highlighted in red). **c–e**, Box plots of methylation levels in oocyte, sperm and zygote (top panels) as well as the distributions of change in methylation levels between sperm and zygote (bottom panels) for specific LINE-1 families, including those that are (c, d) or are not dynamic (e). Top panels: the red line indicates the median, edges the 25th/75th percentile and whiskers the 2.5th/97.5th percentile. Bottom panels: members of each family that are demethylated by greater than 0.45 are highlighted in red. **f–h**, Box plots of methylation levels in oocyte, sperm and zygote (top panels) and the distributions of change in methylation levels between sperm and zygote (bottom panels) for specific families of LTR retroelements, including MMERGLN (f), RLTR10C (g) and IAP elements (h). Top and bottom panels as in c–e. **i**, Mean methylation level for all elements of the L1Md_A LINE (solid blue line) and IAP LTR class (solid red line) that do not markedly change contrasted by LINEs (dashed blue line) and LTR elements (dashed red line) that show the greatest loss at fertilization. SINE elements (green line) are less methylated in sperm than other repeat elements and appear to decrease to oocyte levels.

DMRs also differed substantially in their relative CpG densities (Supplementary Fig. 9).

We next focused specifically on oocyte-contributed promoter DMRs, in part due to their unusual enrichment for high CpG-containing promoters (HCPs). Although this set had no clear functional enrichment, it did include the promoters of several interesting genes that are not expressed during later stages of oogenesis, such as *Dnmt3b* and the somatic isoform of *Dnmt1* (refs 41–43), which suggests a possible regulatory function for at least some of these DMRs. The use of genotyped strains allowed us to confirm that the methylation proximal to the CpG island promoter of copine VII (*Cpne7*), another putative DMR, was directly inherited from the oocyte

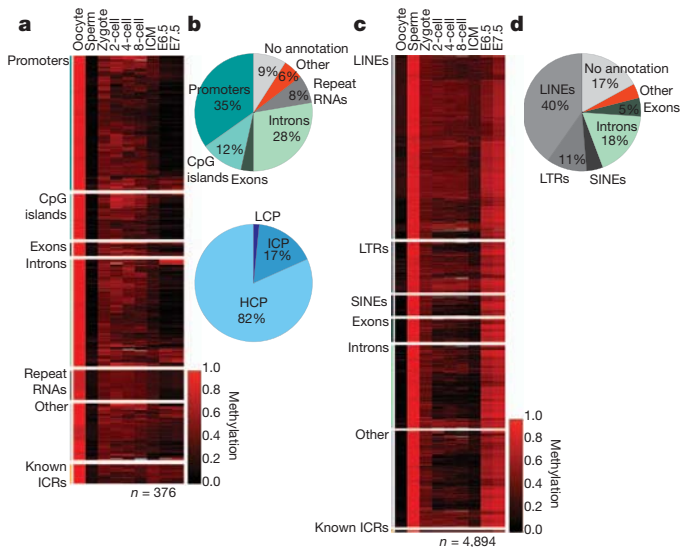


Figure 4 | Differentially methylated regions represent discrete gamete-specific feature classes. **a**, Heat map of methylation levels (black, 0; red, 1; grey, missing value) in 376 identified 100-bp tiles (rows) that behave as oocyte-contributed DMRs in the zygote. Tiles are sorted by functional classes and clustered within each class. Fifteen known ICRs, shown at the bottom, behave similarly in the early embryo and retain intermediate methylation through implantation. Other includes both Other and No annotation. **b**, Genomic features (top) and promoters of different CpG densities (bottom) in oocyte-contributed DMRs. Top: oocyte DMRs are enriched for promoters. Bottom: most of the 105 promoters that overlap oocyte-contributed DMR tiles are high CpG density promoters containing CpG islands (HCPs, light blue). **c**, Heat map of methylation levels (black, 0; red, 1; grey, missing value) in 4,894 identified 100-bp tiles (rows) that behave as sperm-contributed DMRs in pre-implantation embryos. Tiles are sorted and highlighted as in a. **d**, Genomic features in sperm-contributed DMRs are generally intergenic.

(Fig. 5a). As a set, oocyte-contributed promoter DMRs retained intermediate methylation values from the zygote through to the ICM, followed by resolution to hypomethylation in the specified embryo (Fig. 5b, c). Thus, CpG island methylation is transiently stabilized during cleavage divisions before re-establishing an unmethylated state around implantation. A distinct methylation pattern during pre-implantation development is also observed in sperm-contributed DMRs, which retain intermediate methylation values through to the ICM, before being hypermethylated post-implantation to typical somatic levels (Fig. 5d).

RRBS is designed to enrich for CpG dinucleotides (sixfold), but it also captures the other three non-CpG dinucleotides at normal frequencies. Of these, CpA is the predominant target for methylation in mouse and human^{44,45}. Consistent with previous locus-specific observations^{46,47}, we found that oocytes had the highest global CpA methylation level observed across pre-implantation development, and that this level decreased by $\sim 50\%$ in the zygote stage (Fig. 5b and Supplementary Fig. 10). This indicates that non-CpG methylation is inherited as part of oocyte-contributed methylated alleles but is likely lost more rapidly.

Discussion

To better understand the regulation of methylation patterns during its most dynamic phase, we generated genome-scale maps of DNA methylation in both gametes and through implantation. We find that methylation contributed by sperm to the zygote is altered most within retroelements of specific families, whereas other elements remain more protected and retain higher methylation levels throughout development (Supplementary Fig. 11). The methylation status of the oocyte is a strong predictor of levels in the zygote, and regions that are already hypomethylated in the oocyte could explain much of the disparity between the early embryo and sperm. The mechanism and targets of DNA demethylation

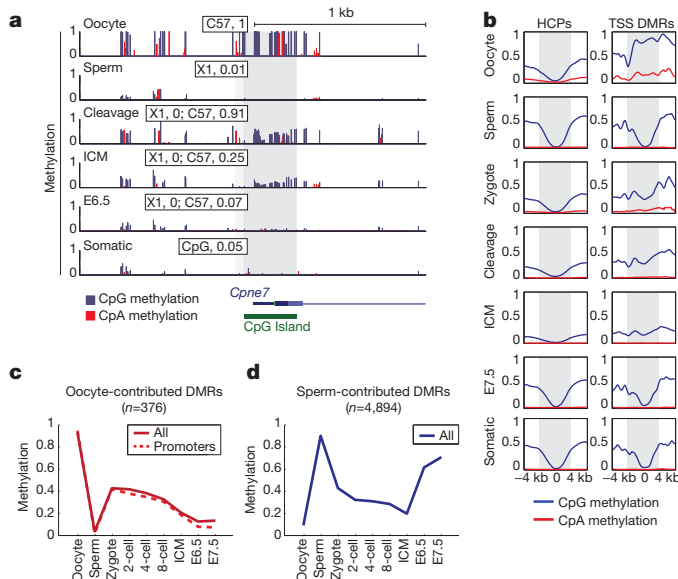


Figure 5 | DMRs resolve after cleavage to univalent hyper- or hypomethylated values in a gamete-of-origin-specific fashion. **a**, Single CpG resolution methylation within 2 kilobases (kb) of the *Cpne7* promoter in gametes and across embryonic development (rows). Dark grey bar highlights the CpG island. A CpG proximal to the island can be tracked to a phase resolving SNP and is highlighted in light grey, with paternal (X1) and maternal (C57) methylation values included as an inset for each trackable phase. Values for SNP methylation in 'cleavage' correspond exactly to those captured in the zygote. Blue bars, CpG methylation; red bars, CpA methylation. **b**, Composite plot of CpG (blue) and CpA (red) methylation for all HCPs (left) and for promoters that are specifically hypermethylated in oocytes (transcription start site (TSS) DMRs, right). The region ± 2 kb of the TSS is marked in grey. Identified promoter DMRs contributed by the oocyte are hypermethylated around the periphery of the TSS and resolve to intermediate values throughout cleavage. An expected HCP methylation architecture is re-acquired for these DMRs around implantation. **c**, Mean methylation levels across stages for oocyte-contributed DMRs in promoters (red, dashed line) versus the complete set (red, solid line). **d**, Sperm-contributed DMRs (blue line) generally resolve to hypermethylation.

during female gametogenesis could possibly be similar to those at work during fertilization³⁴. Regardless, the embryonic pattern more closely resembles that of the oocyte until the later stages of pre-implantation, where DNA methylation is further decreased.

In addition to classical ICRs, which exhibit parent-of-origin-specific methylation maintained through adulthood, a substantial number of CpG island promoters are specifically hypermethylated in the oocyte, in agreement with a recent study²⁵. Surprisingly, these regions retain intermediate values indicating differential allelic methylation before gradually decreasing through ICM specification and gastrulation, where somatic methylation patterns are re-established (Supplementary Fig. 11).

It remains to be investigated whether the diverse targets that exhibit low methylation levels during embryogenesis are the consequence of a single regulatory principle. LINE and LTR activity in the early embryo is associated with some of the earliest transcriptional events during zygotic genome activation. Targeted depletion by antisense oligonucleotides of the L1Md_T class as well as certain LTRs suggests a general requirement for retrotransposon transcription for progression through cleavage^{48,49}. These observations may also support data indicating that the elongation factor/histone acetyltransferase ELP3 is a component of the DNA demethylation machinery and could explain a tight relationship between complete demethylation and transcription-associated complexes⁵⁰.

It is likely that current interest in hmC will spur technical improvements that will permit quantitative dissection of mC and hmC patterns, which will help answer remaining questions regarding

Tet3's universal necessity for conversion to unmethylated cytosines, as well as the effect hmC may have on Dnmt-mediated inheritance³². Tet3's global conversion to hmC of the paternal genome does not seem to lead to equivalently marked demethylation on the basis of the retention of bisulphite-detected signal. The feature-specific dynamics of DNA methylation at fertilization indicate that Tet3 and hmC may be required for targeted demethylation, as well as for driving a gradual depletion through cleavage^{9,10}. Further experiments will be required to characterize this division-dependent demethylation in more detail, and expand it to regions with lower G+C content that are under-represented in RRBS. Notably, other mechanisms must retain heritable methylation information because many targets display relative epigenetic stability from zygote onward and some of these features exhibit embryogenesis-specific methylation patterns.

Our genome-scale single-base-resolution data provide improved understanding of the relationship between DNA methylation and early development. This expands earlier models derived from immunohistochemistry-based observations and begins to address remaining unanswered questions, setting the stage for future epigenetic studies in early mouse development.

METHODS SUMMARY

Gametes, cultured cleavage stage embryos, immunosurgically dissected E3.5 ICM, and post-implantation embryos were isolated as described previously (see Methods). Samples were purified through sequential KSOM microdrops (Millipore) and polar body contaminants mechanically dissected using XY laser (HamiltonThorne) assisted biopsy (Supplementary Fig. 1 and Supplementary Movie 1). Reduced representation bisulphite libraries were generated from proteinase-K-purified, MspI-digested genomic DNA and sequenced using the Illumina Genome Analyzer II platform. Sequenced reads were aligned to the Mouse Genome Build 37 (mm9) using a custom computational pipeline that accounted for the strain identity of samples, which were either inbred or hybrid strains to provide adequate SNP tracking. Sampled cytosines covered $\geq 10\times$ were used for single CpG analysis. Alternatively, single CpGs were incorporated into features taken from ref. 4 or into 100-bp tiles using a $5\times$ threshold. Methylation levels reported for a sample is the average across replicates that met these threshold criteria. Tiles were considered to show a change between two stages if they exhibited a methylation difference ≥ 0.2 and statistical significance through a *t*-test after correction for multiple hypothesis testing (FDR < 0.05) using the Benjamini-Hochberg method. Retrotransposon annotations are from the RepeatMasker track of the UCSC genome browser. Novel DMRs were identified from a pool of 100-bp tiles where one gamete had a mean methylation ≥ 0.75 and the other had a mean methylation ≤ 0.25 . Linear regression applied to this set identified tiles that had zygotic methylation values that fell halfway between those of oocyte and sperm. SNPs between 129X1/SvJ paternal and BDF1 (C57BL/6 x DBA/2) maternal genomes were taken from Mouse Genome Informatics and used to assess relative maternal contamination as well as to identify the parent of origin in order to track allelic methylation values in DMRs and sites exhibiting demethylation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 July 2011; accepted 15 February 2012.

Published online 28 March 2012.

- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (Suppl.), 245–254 (2003).
- Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* **9**, 465–476 (2008).
- Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
- Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).
- Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
- Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).

8. Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2**, 241 (2011).
9. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
10. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194 (2011).
11. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).
12. Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* **6**, 705–714 (1992).
13. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42**, 1093–1100 (2010).
14. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nature Biotechnol.* **28**, 1079–1088 (2010).
15. Razin, A. & Shemer, R. DNA methylation in early development. *Hum. Mol. Genet.* **4**, 1751–1755 (1995).
16. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382 (1987).
17. Rougier, N. *et al.* Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev.* **12**, 2108–2113 (1998).
18. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
19. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88–93 (2003).
20. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
21. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241**, 172–182 (2002).
22. Kim, S. H. *et al.* Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos. *Biochem. Biophys. Res. Commun.* **324**, 58–63 (2004).
23. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnol.* **28**, 1106–1114 (2010).
24. Harris, R. A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature Biotechnol.* **28**, 1097–1105 (2010).
25. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43**, 811–814 (2011).
26. Davis, T. & Vaisvila, R. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *J. Vis. Exp.* **48** (2011).
27. Fic, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
28. Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
29. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
30. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25**, 679–684 (2011).
31. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**, 451–464 (2011).
32. Branco, M. R., Fic, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Rev. Genet.* **13**, 7–13 (2011).
33. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241**, 172–182 (2002).
34. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78–82 (2010).
35. Wossidlo, M. *et al.* Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
36. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
37. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
38. Ichiyanagi, K. *et al.* Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. *Genome Res.* **21**, 2058–2066 (2011).
39. Goodier, J. L., Ostertag, E. M., Du, K. & Kazazian, H. H. Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* **11**, 1677–1685 (2001).
40. Edwards, C. A. & Ferguson-Smith, A. Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **19**, 281–289 (2007).
41. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9**, 2395–2402 (2000).
42. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* **22**, 1607–1616 (2008).
43. Lucifero, D. *et al.* Coordinate regulation of DNA methyltransferase expression during oogenesis. *BMC Dev. Biol.* **7**, 36 (2007).
44. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl Acad. Sci. USA* **97**, 5237–5242 (2000).
45. Ziller, M. J. *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7**, e1002389 (2011).
46. Haines, T., Rodenhiser, D. & Ainsworth, P. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Dev. Biol.* **240**, 585–598 (2001).
47. Tomizawa, S. *et al.* Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**, 811–820 (2011).
48. Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E. & Spadafora, C. Expression of LINE-1 retrotransposons is essential for murine preimplantation development. *Mol. Reprod. Dev.* **73**, 279–287 (2006).
49. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 (2003).
50. Okada, Y., Yamagata, K., Hong, K., Wakayama, T. & Zhang, Y. A role for the elongator complex in zygotic paternal genome demethylation. *Nature* **463**, 554–558 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank all members of the Meissner and Regev laboratories. M. Garber, N. Yosef, J. Ye, R. Koche, C. Bock, R. Maehr and D. Egli for technical advice and discussion. We thank all members of the Broad Sequencing Platform, in particular F. Kelly and J. Meldrim, T. Fennel, K. Tibbetts and J. Fostel. We also thank S. Levine, M. Gravina and K. Thai from the MIT BioMicro Center. A.R. is an investigator of the Merkin Foundation for Stem Cell Research at the Broad Institute. This work was supported by the NIH Pioneer Award (5DP1OD003958), the Burroughs Wellcome Career Award at the Scientific Interface and HHMI (to A.R.), the Harvard Stem Cell Institute (to T.S.M.) and the NIH (5RC1AA019317, U01ES017155 and P01GM099117), the Massachusetts Life Science Center and the Pew Charitable Trusts (to A.M.) and a Center for Excellence in Genome Science from the NHGRI (1P50HG006193-01, to A.R. and A.M.).

Author Contributions Z.D.S. and A.M. conceived the study and Z.D.S., M.M.C. and A.M. facilitated its design. Z.D.S. collected samples and performed methylation profiling. M.M.C. performed all analysis with assistance from T.S.M. and Z.D.S. H.G. and A.G. provided critical technical assistance and expertise. Z.D.S., M.M.C., T.S.M., A.R. and A.M. interpreted the data. Z.D.S., M.M.C. and A.M. wrote the paper with the assistance of the other authors.

Author Information RRBS data is deposited at the Gene Expression Omnibus under accession number GSE34864. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M. (alexander_meissner@harvard.edu).

METHODS

Preparation of samples. Isolation of gametes, pre- and post-implantation embryos was performed using procedures described in detail elsewhere⁵¹. Briefly, 4–6-week-old BDF1 female mice (Charles River) were injected with 5 IU of pregnant mare gonadotropin (Sigma) followed 46 h later by 5 IU human chorionic gonadotropin (Sigma). Primed mice were then either directly used to collect oocytes or mated with 129X1 male mice (Jackson) to collect fertilized embryos. Twelve hours after final hormone injection, oocytes or zygotes were isolated from the ampulla under mineral oil and collected in hyaluronidase-containing M2 medium (Millipore) drops to eliminate cumulus cells or sperm contaminants. Oocytes were then depleted of somatic contaminants via progressive dilution through sequential drops of CO₂ buffered, amino-acid-supplemented KSOM medium (Millipore) until no somatic contaminants were observed.

Embryos were cultured in KSOM until collection at progressive cleavage stages with isolation occurring within 6 h of the first observed cleavage event for that stage. Zygotes were screened for the presence of visible pronuclei and subjected to XY Clone (Hamilton Thorne) laser-assisted polar body biopsy using an 8- μ m bore piezo pipette (Humagen, Supplementary Fig. 1 and Supplementary Movie 1). Clean cleavage stage embryos underwent an identical approach, with developmental progression unhindered by biopsy conducted at the 2-cell stage (Supplementary Fig. 1). For each collection, batches of embryos were carefully screened to ensure that each stage did not contain any abnormal embryos. Collection for zygotes was timed at ~10 h.p.f. with fertilization assumed to occur 6–8 h after HCG injection, which was again confirmed by the relative synchronicity of the first cleavage division and by relative pronuclear stage. Biopsies were conducted in M2 media (Millipore) in batches of 5–10 embryos to reduce time on the micro-manipulator stage. Before the final collection, cleaned and sorted samples were washed with acid Tyrode's solution (Sigma) to eliminate the zona pellucida and to deplete any residual somatic contaminants or polar bodies through a short series of additional washes.

Cavitated E3.5 blastocysts were flushed from the uteri of naturally mated mice into M2 or DMEM followed by sequential washing in KSOM. The ICM itself was enriched from collected blastocysts by treating the embryo with rabbit anti-mouse serum (Sigma) before immunosurgical depletion of the trophectoderm using guinea-pig complement serum (Sigma). Isolated ICMs were serially washed to remove contaminants.

E6.5 and 7.5 embryos were isolated after mechanical dissection of the decidua from the uterine lining of mated mice. Samples were again serially washed and peripheral trophectodermal tissues dissected away using fine glass capillaries.

Swimming sperm samples were isolated in BSA-supplemented human tubule fluid (Millipore) from the caudal epididymis of male mice within 5 days of a successful natural mating as scored by copulation plug. Sperm were incubated in buffered HTF as in *in vitro* fertilization for over 1 h in part to reduce somatic contaminants, and samples were scored for relative quality under $\times 10$ microscopy before snap freezing.

All samples were then collected at minimal volume and either snap frozen or immediately re-suspended in DNA lysis buffer.

Preparation of reduced representation bisulphite sequencing libraries. RRBS libraries were generated as described previously^{52–54}. Briefly, DNA was isolated from snap-frozen embryos in DNA lysis buffer (100 mM Tris-HCl (pH 8.5), 5 mM EDTA, 0.2% SDS, 200 mM NaCl) supplemented with 300 μ g ml⁻¹ proteinase K (Invitrogen) followed by phenol:chloroform extraction, ethanol precipitation and re-suspension in EB buffer. Isolated DNA was then subjected to MspI digestion (NEB), end repair using Klenow 3'-5' exo- (NEB) supplemented with GTP, meCTP and ATP in a 1:1:10 ratio to facilitate 3' A tailing, and ligation of standard adapters using ultraconcentrated 10⁶ U T4 DNA ligase (NEB) and extended 20 h ligation at 16 °C. Size selection of 40–150 and 150–270-bp fragments containing ligated adaptor was conducted by extended gel electrophoresis using NuSieve 3:1 agarose (Lonza) and gel extraction (Qiagen) using 50 ng dephosphorylated, sonicated *Escherichia coli* DNA as a protective carrier and to increase overall yield. The isolated molecular weight fractions in a given RRBS library were then separately treated with sodium bisulphite using the EpiTect bisulphite conversion and column purification system (Qiagen) with a modified conversion strategy as described⁵². After clean up, the optimal, minimum PCR cycle number required to generate the final libraries was gauged using diagnostic PCRs for each library. Final libraries were then generated from the complete bisulphite converted pool and purified through a second round of gel electrophoresis. High- and low-molecular-mass fragments were then either sequenced separately or pooled at a 2:1 ratio by mass to assume an equimolar representation of both size ranges. Libraries were then sequenced on an Illumina Genome Analyzer II before alignment and analysis. The sequencing reads were aligned to the Mouse Genome Build 37 (mm9) using a custom computational pipeline taking into account the

strain background for each sample^{4,54}. To supplement our data set we included sperm replicate 2 from ref. 25 (SRA accession ERP000689).

Estimating methylation levels. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T. Single CpG methylation levels were limited to those CpGs that had at least tenfold coverage. For 100-bp tiles, reads for all the CpGs that were covered more than fivefold within the tile were pooled and used to estimate the methylation level as described for single CpGs. The CpG density for a given single CpG is the number of CpGs 50 bp up- and downstream of that CpG. The CpG density for a 100-bp tile is the average of the CpG density for all single CpGs used to estimate methylation level in the tile. CpA methylation levels were estimated in the same way as for CpG methylation.

The methylation level reported for a sample is the average methylation level across replicates. A replicate will contribute to the average only if it meets the coverage criteria within the replicate. Technical replicates were averaged before contributing to the sample average.

Genomic features. High-density CpG promoters (HCP), intermediate-density CpG promoters (ICP), low-density CpG promoters (LCP), TSS, CpG island, and DMR annotations were taken from ref. 4. Promoters are defined as 1 kb up- and downstream of the TSS. LINE, LTR and SINE annotations were downloaded from the UCSC browser (mm9) RepeatMasker tracks. Gene annotations were downloaded from the UCSC browser (mm9) RefSeq track. In each case, the methylation level of an individual feature is estimated by pooling read counts for all CpGs within the feature that are covered greater than fivefold, and levels are only reported if a feature contains at least 5 CpGs with such coverage (in contrast to 100-bp tiles where no minimum number of CpGs is required). A tile is annotated as a genomic feature if any portion of the tile overlaps with the feature and may be annotated by more than one feature (for example, the same tile can be annotated as both a promoter and a gene).

Identification of tiles with changing methylation levels and their enrichments. A tile is considered changing if it both has a methylation difference ≥ 0.2 between two stages and is significant in a two-sample *t*-test with unequal variance after correction for multiple hypothesis testing (FDR < 0.05) using the Benjamini-Hochberg method⁵⁵. Enrichment *P* values are from the hypergeometric distribution where the background is the number of tiles that have a methylation difference ≥ 0.2 and are corrected for multiple hypotheses at FDR < 0.05 , based on the number of feature sets tested.

Identification of enriched retrotransposon families. The same procedure for identifying changing tiles was applied to the methylation levels of retrotransposon elements to identify changing elements. Enrichment for families was done using annotations from the RepeatMasker track of the UCSC genome browser.

Novel DMR identification. 100-bp tiles where one gamete had a mean methylation greater than 0.75 and the other gamete had a mean methylation of less than 0.25 were flagged as potential DMRs. Linear regression was used to identify tiles which had methylation levels in zygote which were halfway between the methylation levels in oocyte and sperm. Only tiles that had two replicates present in each time point were considered. Residuals were calculated as the mean of the differences between the model predictions and the data taking into account missing values. ANOVA was used to assign a *P* value to each tile. A tile was considered a novel DMR if it had a residual in the tenth percentile of tiles tested and a significant *P* value from ANOVA with a Benjamini-Hochberg FDR < 0.05 . A residual in the tenth percentile corresponds to an FDR < 0.1 by a permutation test where zygote methylation values are shuffled for potential DMR tiles. In the pie charts (Fig. 5b, d), the genomic feature that covered the most novel tiles was reported first and then subtracted from the set before reporting the feature which covered the next largest number of tiles. This procedure was repeated until all tiles were categorized. The one exception was for oocyte-contributed DMRs where promoters were taken out before genes.

Identification of SNPs. An initial set of SNPs between 129X1 and BDF1 (C57BL/6 \times DBA/2) was taken from Mouse Genome Informatics⁵⁶. The set was filtered such that SNPs that fell into the following categories were removed: (1) SNPs that had inconsistent entries for the same position; (2) SNPs not trackable by RRBS (C/T or A/G); (3) SNPs between C57BL/6 and DBA/2; and (4) SNPs that were not covered by X1 and BDF1 in an *in silico* digest. The log odds ratio [$\log_2(X1 \text{ count} + 0.01 / C57 \text{ count} + 0.01)$] was calculated for each SNP that was covered in the data set ($n = 786$). SNPs that had at least 10 \times coverage with an average log odds ratio across all replicates between -5 and 5 and a sperm X1 log odds ratio greater than 2 were considered to be of stringent quality ($n = 636$) and used to assess both maternal bias and to serve as a general quality control metric for all libraries incorporated into the data set.

Parent-of-origin methylation tracking. The 636 SNPs identified above corresponded to 1,674 CpG dinucleotides and were used to track allelic single CpG methylation. Reads were segregated into either X1 or BDF1 backgrounds according

to SNP type, and CpG methylation levels were called in the same manner described above. SNP normalized methylation values (Supplementary Fig. 4) are the average of the methylation values from each strain.

51. Nagy, A. *Manipulating the Mouse Embryo: a Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2003).
52. Smith, Z. D., Gu, H., Bock, C., Gnirke, A. & Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**, 226–232 (2009).
53. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols* **6**, 468–481 (2011).
54. Gu, H. *et al.* Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nature Methods* **7**, 133–136 (2010).
55. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
56. Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E. & Eppig, J. T. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.* **39**, D842–D848 (2011).

The 2.8 Å crystal structure of the dynein motor domain

Takahide Kon^{1,2}, Takuji Oyama¹, Rieko Shimo-Kon¹, Kenji Imamula³, Tomohiro Shima⁴, Kazuo Sutoh⁵ & Genji Kurisu^{1,2}

Dyneins are microtubule-based AAA⁺ motor complexes that power ciliary beating, cell division, cell migration and intracellular transport. Here we report the most complete structure obtained so far, to our knowledge, of the 380-kDa motor domain of *Dictyostelium discoideum* cytoplasmic dynein at 2.8 Å resolution; the data are reliable enough to discuss the structure and mechanism at the level of individual amino acid residues. Features that can be clearly visualized at this resolution include the coordination of ADP in each of four distinct nucleotide-binding sites in the ring-shaped AAA⁺ ATPase unit, a newly identified interaction interface between the ring and mechanical linker, and junctional structures between the ring and microtubule-binding stalk, all of which should be critical for the mechanism of dynein motility. We also identify a long-range allosteric communication pathway between the primary ATPase and the microtubule-binding sites. Our work provides a framework for understanding the mechanism of dynein-based motility.

Dyneins are large microtubule-based motor complexes responsible for various biological movements¹. Cytoplasmic dyneins power diverse cellular processes, such as cell division, cell migration and minus-end-directed intracellular transport of vesicles and organelles along microtubules^{2–4}. Axonemal dyneins are specialized to drive ciliary and flagellar beating^{5,6}.

Unlike the other families of cytoskeletal motor proteins—myosin and kinesin, the heavy chains of which are evolutionarily close to the G protein families—the dynein heavy chains (molecular mass ≥ 500 kDa) belong to the AAA⁺ superfamily of mechanochemical enzymes⁷. Members of this ATPase family participate in diverse cellular functions including protein degradation, protein-complex disassembly, and DNA replication and recombination⁸. Nevertheless, AAA⁺ proteins share structural features: they contain conserved ATPase units called AAA⁺ modules, and they operate as oligomers in which typically six AAA⁺ modules assemble into a ring structure⁹. Although in dynein six distinct AAA⁺ modules (AAA1–AAA6) are concatenated into a single polypeptide⁷, their overall arrangement is similar to that observed in other ring-shaped AAA⁺ proteins^{10–12}. Therefore, dynein probably shares common mechanistic features with AAA⁺ ring ATPases rather than with the G-protein-related motors. In addition to the ATP-hydrolysing AAA⁺ ring, the dynein heavy chain contains two major functional units: the stalk and linker. The stalk has an ATP-sensitive microtubule-binding site (MTBD) at its tip^{13,14}, whereas the linker has been proposed to function as a mechanical element for generating dynein's power stroke^{10,12,15}.

The molecular mechanism of action of dynein is still poorly understood compared with that of other cytoskeletal motors and major AAA⁺ proteins, which is partly due to the lack of high-resolution structural information. Most recently, two crystal structures of the motor domain of cytoplasmic dynein were reported at 4.5–6 Å resolution^{16,17}, revealing the dynein structures at the secondary structure level. However, a much higher-resolution analysis is required to reveal the complete atomic structure of the motor domain, which is critical for understanding the structural basis of dynein's force generation and movement. Here we report the crystal structure of the motor domain of *Dictyostelium* cytoplasmic dynein at 2.8 Å resolution, with approximately 90% of its amino acid residues traced.

Overall structure of dynein motor domain

We crystallized the wild type and a truncation mutant (Δ MTBD) of the 380-kDa motor domain (residues V1388–I4730 of the heavy chain) of *Dictyostelium discoideum* cytoplasmic dynein^{18,19} (Fig. 1a). Δ MTBD lacks the small MTBD (residues A3372–K3495), but retains ATPase activity ($111 \pm 6 \text{ s}^{-1}$) comparable to the maximal rate (k_{cat}) of microtubule-activated ATPase in the wild-type motor domain²⁰. For Δ MTBD, the structure was solved using three heavy-atom derivatives. The resulting electron-density maps were of sufficient quality (Supplementary Fig. 1a–f) to create a nearly complete atomic model (Fig. 1b), except for the 22 amino-terminal residues, disordered loops, and truncated MTBD (Supplementary Methods). The model was refined to 2.8 Å resolution ($R_{\text{work}}/R_{\text{free}}$ of 26.3/32.1%; Supplementary Table 1). For wild-type dynein motor domain, the structure was solved by molecular replacement using the Δ MTBD structure as a search model, and the model including MTBD was refined to 3.8 Å resolution ($R_{\text{work}}/R_{\text{free}}$ of 22.1/29.4%; Fig. 1c and Supplementary Table 1). In both structures, the crystallographic asymmetric unit has two motor molecules (Δ MTBD: Δ MTBD-A and Δ MTBD-B; wild type: WT-A and WT-B; Supplementary Fig. 1g, h). The four molecules are similar in overall structure, but have several distinct differences as described below, indicative of dynamic structural changes of the motor domain. Here we describe the monomeric motor domain structure by referring mainly to Δ MTBD-A, because it provides the most complete model.

The motor domain comprises the AAA⁺ ring and three additional structural units: the linker, stalk/strut and C sequence (Fig. 1b, c and Supplementary Fig. 2). The central ATP-hydrolysing ring has six AAA⁺ modules arranged around the ring central pore, and is 130 Å in diameter, 65 Å in thickness and 35 Å in the pore diameter (Supplementary Fig. 3a). Each AAA⁺ module contains the N-terminal α/β -submodule and the carboxy-terminal α -submodule (Supplementary Fig. 4) that constitute the front and back faces of the ring, respectively (Supplementary Fig. 3a). The α/β -submodule core exhibits a Rossmann-type fold (Supplementary Fig. 4a), which is commonly found in nucleotide-binding proteins^{21,22}. AAA1–AAA4 have unique insertions in their α/β -submodules: H2, pre-Walker B and pre-sensor

¹Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. ²Department of Macromolecular Science, Graduate School of Science, Osaka University, Osaka 560-0043, Japan. ³Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan. ⁴Department of Physics, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁵Research Institute for Science and Engineering, Waseda University, Takada 1-17-22, Toshima-ku, Tokyo 171-0033, Japan.

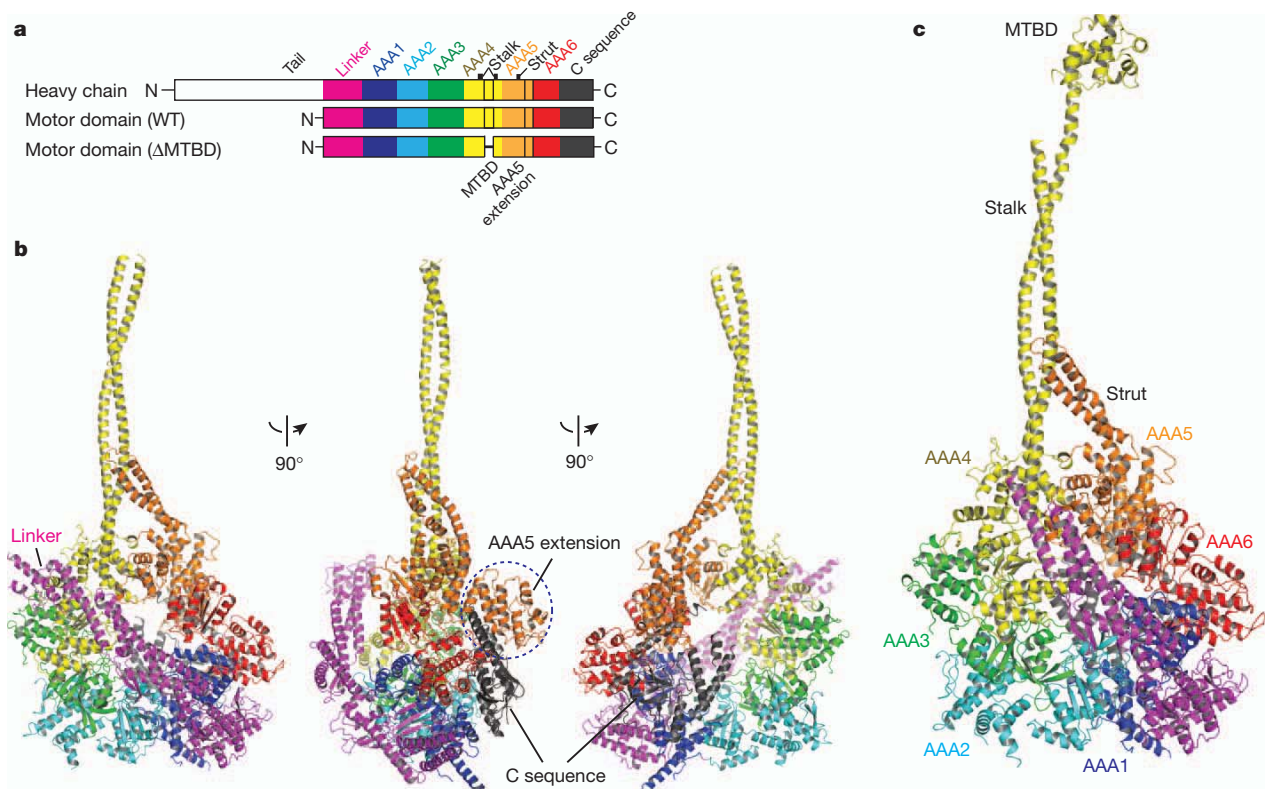


Figure 1 | Overall structure of the dynein motor domain. **a**, Sequence diagrams of dynein heavy chain and wild-type and Δ MTBD motor domains, showing tail, linker and six AAA⁺ modules constituting ring, stalk-strut and

C-sequence. The colour code shown here is used for all figures, unless otherwise noted. **b**, **c**, Ribbon representation of Δ MTBD (**b**) and wild-type (**c**) motor domain structures.

I (PS-I) inserts (Supplementary Fig. 4b). AAA5 contains an additional globular structure (named here the AAA5 extension) that contains eight α -helices (H9–H16) and protrudes from the back face of the ring (Supplementary Fig. 3a).

Above the front face of the ring, the linker exists as a rod-like structure that comprises nineteen α -helices and eight β -strands and can be partitioned into five subdomains from 0 to 4 (Supplementary Fig. 3b). This structure extends from AAA1- α/β above the ring to the periphery of AAA4 (Fig. 1b, c). Between AAA4 and AAA5, two prominent coiled coils protrude and form a Y-shaped structure (Fig. 1b, c and Supplementary Fig. 3c). The longer coiled coil extending from AAA4 is the stalk and the shorter one extending from AAA5 is the strut (the buttress) that was identified in previous X-ray studies^{16,17}. The WT-A stalk includes MTBD at the tip, whereas in the WT-B stalk, MTBD and its distal portion (K3359–K3512) are missing, probably due to structural flexibility of the stalk. On the ring's back face, the C sequence exists as a complex structure comprising six α -helices and an incomplete six-stranded antiparallel β -barrel (Supplementary Fig. 3d). The shape of the C sequence is distinctively flat, spreading over the α -submodules of AAA1, AAA5 and AAA6 (Fig. 1b).

This overall structure is similar at the secondary structure level to the medium-resolution structures^{16,17} (see Supplementary Fig. 5 for comparison with our previous 4.5 Å model, which corrects errors in assignment in the previous study¹⁷). However, unlike previous models, our structure is reliable enough to discuss the dynein structure and mechanism at the level of individual amino acid residues, as judged from the reasonable refinement statistics at 2.8 Å resolution and resulting electron-density maps.

The four nucleotide-binding sites

The motor domain has four ADP molecules bound to the first four AAA⁺ modules (Fig. 2a), in agreement with a previous biochemical study²³. Each nucleotide-binding site is formed at the interface

between three submodules, α/β and α of one module and α/β of the adjacent module (Fig. 2b–e and Supplementary Fig. 6), with a structural organization very similar to that of other AAA⁺ ATPases^{9,21,22}. At the AAA1 nucleotide-binding site (Fig. 2b), the phosphate moiety of ADP is surrounded by the Walker-A motif (G1977–T1981) in AAA1- α/β , whereas the ribose moiety is recognized by the sensor-II motif (R2150) and neighbouring residues (L2149 and K2153) in AAA1- α . The adenine base is accommodated by hydrophobic residues in AAA1- α (I2108) and the N loop (L1947 and V1948) connecting the linker with the AAA1 ATPase site (Fig. 2a, b and Supplementary Fig. 6a). Given that ATP is the only energy source for dynein motor function²⁴ although it can hydrolyse other nucleotides²⁵, the adenine base recognition by the N loop is probably involved in the coupling of the linker actions and ATPase. Residues in Walker-B (D2026 and E2027) and sensor-I (N2078) motifs of AAA1- α/β , which are generally required for ATP hydrolysis, occupy the nucleotide-binding site. In addition, an arginine finger (R2410) crucial for ATP hydrolysis stretches from the adjacent AAA2- α/β and points to the bound ADP.

Among the AAA1–AAA4 nucleotide-binding sites with similar structural frameworks, AAA2 shows several distinct characteristics (Fig. 2c). First, the crucial sensor-II motif is missing and L2531 occupies the position. Second, in the Walker-B motif, the conserved glutamic acid essential for ATP hydrolysis is replaced by glycine (G2357). Moreover, two arginine fingers (R2806 and R2809) from the adjacent AAA3 module interact extensively with the phosphates of bound ADP. The unique nucleotide-binding mode of AAA2 indicates that this site binds ADP and/or ATP but does not hydrolyse ATP. In contrast to AAA2, most functional motifs of AAA3 and AAA4 required for ATP binding and hydrolysis are oriented in a canonical manner (Fig. 2d, e), indicating that dynein has three ATP hydrolysis sites—AAA1, AAA3 and AAA4. This was further supported by pre-steady-state kinetic analysis of phosphate release from the motor domain carrying multiple mutations in the three nucleotide-binding sites (Supplementary Fig. 7

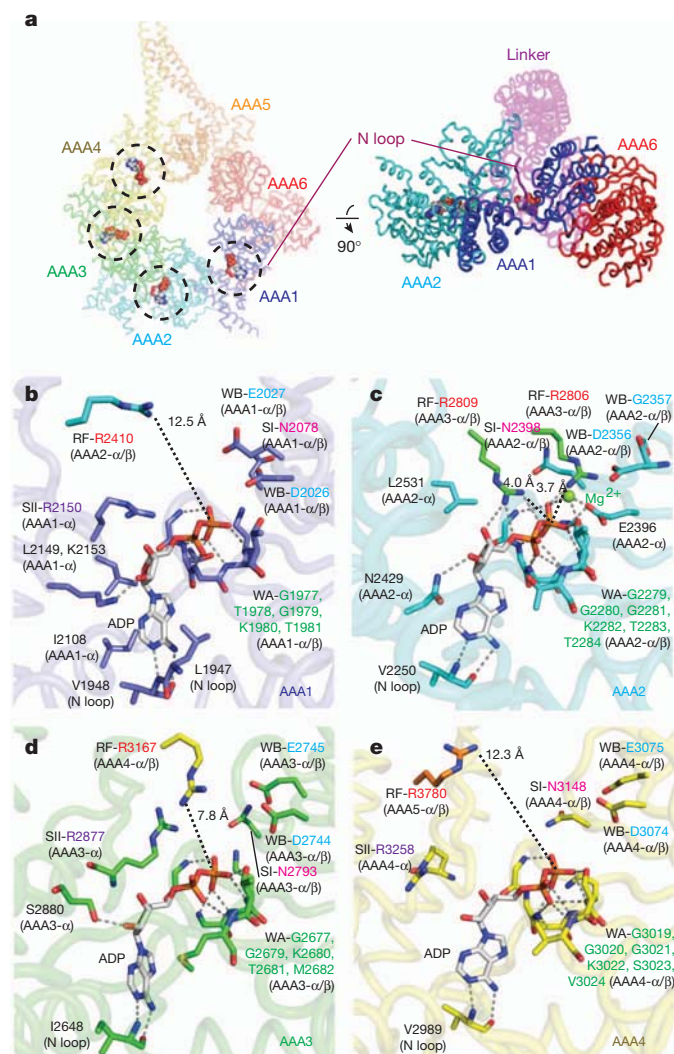


Figure 2 | Structure of the four nucleotide-binding sites. **a**, Locations of bound ADP in the ring. ADP molecules are shown as space-filled models. The right panel shows the bottom view of the ring and linker. The N loop connects the AAA1 nucleotide-binding site and linker. **b–e**, Magnified views of ADP in the nucleotide-binding sites. ATPase motifs, that is, Walker-A, Walker-B, sensor-I, sensor-II and arginine-finger, are abbreviated as WA, WB, SI, SII and RF, respectively. Residues that are hydrogen-bonded to ADP are shown. In panel **b**, residues participating in ribose moiety (L2149) and adenine ring (I2108 and L1947) accommodation are included. In panel **c**, L2531 occupying the AAA2 sensor-II position is shown. Other residues forming hydrophobic interactions with ADP are omitted for clarity. Hydrogen bonds to ADP are drawn with grey dotted lines. Distances between the arginine finger Nη and ADP β-phosphate are indicated by black dashed lines.

and Supplementary Table 2). Similar kinetic experiments performed with microtubules showed that among the three ATPase sites, AAA1 alone drives microtubule-activated ATPase essential for dynein function (Supplementary Fig. 7 and Supplementary Table 2).

Open and closed ring conformations

The ring structure of ΔMTBD-A displays a unique arrangement of six AAA⁺ modules (upper, Supplementary Fig. 8a). The six α-submodules are almost equally arranged in pseudo-six-fold symmetry. In contrast, the six α/β-submodules exhibit an asymmetrical arrangement and can be grouped into AAA5–AAA6–AAA1, forming one loosely packed half-ring structure with buried inter-modular surfaces of approximately 200 Å², and AAA2–AAA3–AAA4, comprising the more tightly packed half-ring structure with buried inter-modular surfaces of 500–1,000 Å². There is no direct contact between the

two half-rings, and, as a result, apparent gaps are formed between them. This asymmetrical arrangement is different from that previously reported for the yeast motor domain in a no-nucleotide state, in which gaps exist between AAA1–AAA2 and between AAA5–AAA6 instead¹⁶.

The AAA1–AAA2 and AAA4–AAA5 gaps lie in the AAA1 and AAA4 nucleotide-binding sites, respectively, affecting the positioning of the trans-acting arginine finger critical to ATP hydrolysis in many ATPases²⁶. With these open gaps, the arginine fingers extending from AAA2 to AAA1 (R2410) and from AAA5 to AAA4 (R3780) are positioned far away from the bound ADP (Fig. 2b, e and Supplementary Fig. 8a) to allow ATP hydrolysis when bound to these nucleotide-binding sites. Given that ATP is hydrolysed at AAA1 and AAA4, the gaps are expected to close during their ATPase cycles so that arginine fingers can interact with the γ-phosphate of bound ATP. Indeed, the WT-A ring structure has a more closed conformation (lower, Supplementary Fig. 8a). This closure shortens the distance between the arginine finger and bound ADP (Supplementary Fig. 8a, b), which would allow R2410 and R3780 to participate in ATP hydrolysis. Although ΔMTBD and the wild-type dynein motor domain are in the same nucleotide-bound state, the structural differences suggest the existence of an ATPase-dependent open/closed transition of the ring, as observed for other AAA⁺ ATPases during their mechanochemical actions^{27,28}.

Linker and ring interactions

To understand the mechanism of dynein motility, it is critical to elucidate the structural details of interactions between the ATP-hydrolysing ring and the linker whose swing-like motions have been proposed to be the major contributor to dynein's power stroke^{10,12,15}. In the present structure, the linker position is close to that of the post-power-stroke state inferred from electron microscopy analysis¹². When viewed from the ring's front face, the linker appears to bridge AAA1 and AAA4 by spanning the inner edge of AAA2, AAA3 and AAA4 (Fig. 1b, c). However, the linker interacts with the ring only through two major interfaces at AAA1–AAA6 and AAA2 (Fig. 3a). The first interaction interface is at the linker base connecting to AAA1-α/β (orange in Fig. 3a). Here, linker subdomains 3 and 4 are tightly sandwiched between AAA1-α/β, AAA1-α, and AAA6-α through four hydrogen bonds and hydrophobic interactions involving 64 residues. The resulting buried surfaces are sufficiently wide (2,478 Å²) for stable interactions between the linker base and ring. The second interface is at the middle region of the linker (cyan in Fig. 3a). Here, two finger-like structures—the H2 insert β-hairpin and PS-I insert loop—protruding from AAA2 interact with linker subdomains 2 and 3, respectively (Fig. 3b and Supplementary Fig. 9): the H2 insert tip forms hydrophobic interactions with β1, β2 and the H7 helix in linker subdomain 2, and the PS-I insert tip forms a hydrophobic interaction with the H13 helix in linker subdomain 3. Several hydrophobic interactions are formed between H2 and PS-I inserts, indicating that these two finger-like structures function together to support the linker structure.

To examine the functional significance of H2 and PS-I inserts, we performed functional analysis of a mutant (ΔH2I) in which the H2 insert β-hairpin (S2319–D2329) was replaced with a flexible linker (GG). With this mutation, all H2 insert residues interacting with the linker and PS-I insert were deleted or replaced (Fig. 3b). ΔH2I exhibited clear microtubule-activated ATPase activity and ATP-induced change in the affinity for microtubules, although at lower levels than those in wild type (Fig. 3c and Supplementary Fig. 10a). However, steady-state fluorescence resonance energy transfer (FRET)-based analysis¹⁵ did not detect ATPase-dependent linker motions in the mutant (Fig. 3d), indicating the critical involvement of H2/PS-I inserts in linker swing actions. The FRET signal reporting the linker position suggests that the linker may adopt an intermediate position between unprimed (post-power-stroke) and primed (pre-power-stroke) positions regardless of the nucleotide state.

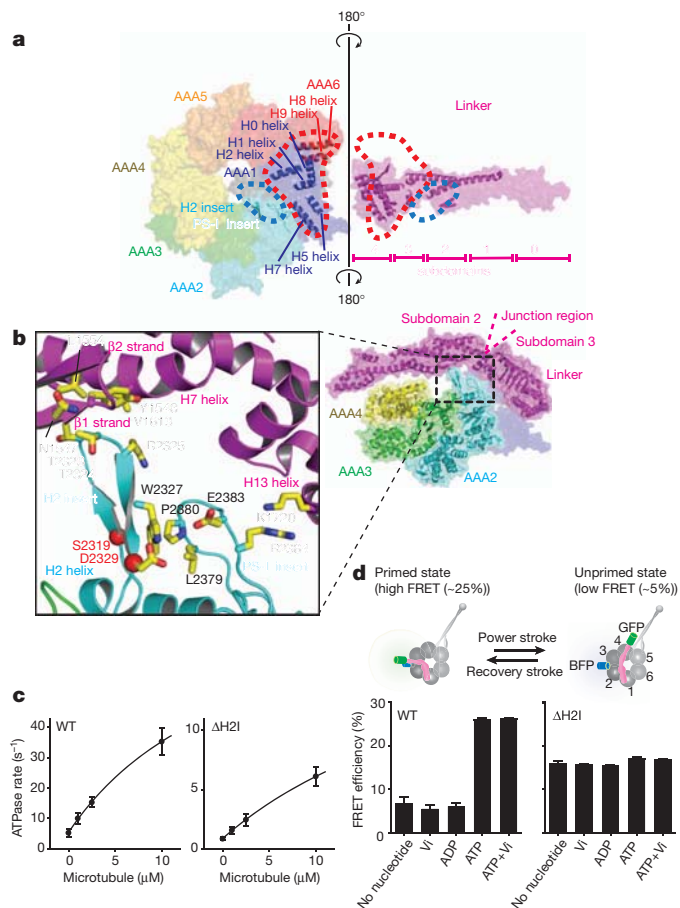


Figure 3 | Linker-ring interactions. **a**, Open-book presentation of linker-ring interactions in Δ MTBD-A. First and second interaction interfaces are surrounded by dotted circles in orange and cyan, respectively. The secondary structure elements involved in the interactions are shown in cartoon representation. **b**, Side view of the motor domain showing linker-AAA2 interactions in Δ MTBD-A. Interacting residues between linker subdomain 2/3 and H2/PS-I inserts and between H2 and PS-I inserts are shown. D2329 and E2383 form a hydrogen bond network through R2337 (not shown). Truncation positions (S2319 and D2329) in Δ H2I are indicated. **c**, Microtubule-activated ATPase rates of the motor domain. Each symbol indicates mean \pm s.d. ($n = 3$). **d**, FRET efficiency between BFP and GFP moieties in the motor domain in the presence of the indicated nucleotide. Vi, vanadate. The high and low FRET of wild type represent linker positions at primed and unprimed states, respectively¹⁵. Mean \pm s.d. ($n = 3$) are shown.

The conventional linker swing model assumes that in the unprimed state, the N-terminal region of the linker docks on AAA4 or AAA5 with its C-terminal base connected to AAA1, and that ATPase-induced structural changes propagate from AAA1 to AAA4–AAA5 to release the docked linker end and induce the linker swing^{10,12,16}. Here we propose an alternative model, on the basis of our findings that there is no direct contact between the linker and AAA3–AAA5 in our structure (Fig. 3b and Supplementary Fig. 9) and that H2 and PS-I inserts of AAA2 have a critical role in linker swing (Fig. 3d). Given that H2 and PS-I inserts in other AAA⁺ proteins serve as interaction interfaces with target proteins or DNA to catch and remodel the targets^{22,29}, dynein could also use this remodelling apparatus for the linker swing. During the mechanochemical cycle of dynein, H2 and PS-I inserts interacting with linker subdomains 2 and 3 could remodel the junction of the two subdomains so that the linker bends there and swings against the ring. In support of the model, the subdomain junction is the thinnest part of the linker (Fig. 3a, b) and therefore is the best candidate for the bending site among subdomain junctions along the linker. The open-closed transition of the AAA1–AAA2 interface during the primary AAA1 ATPase cycle (Supplementary

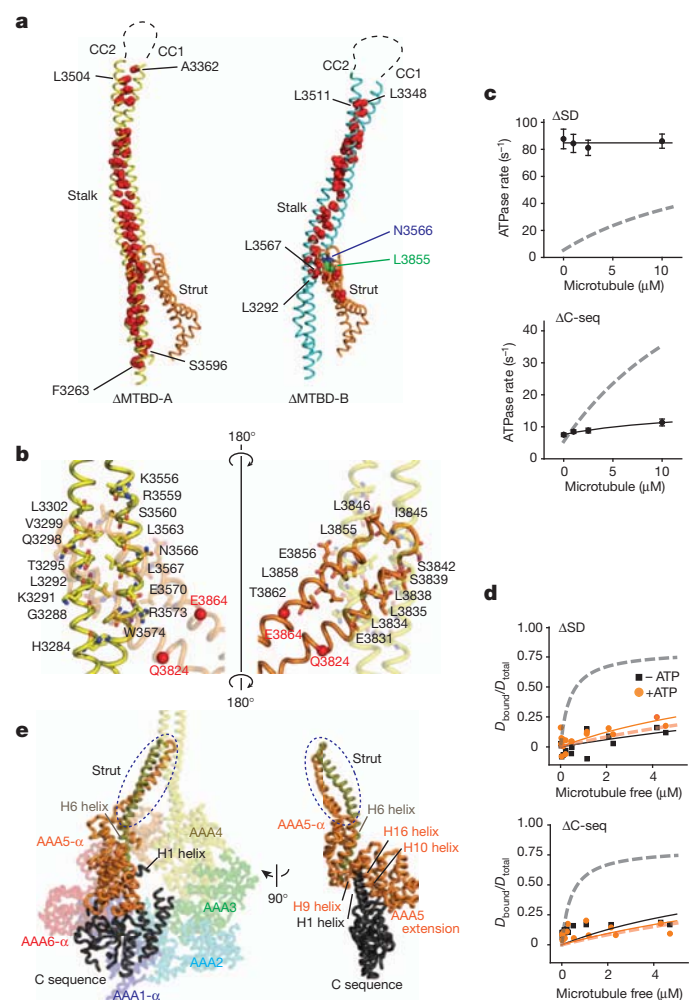


Figure 4 | Structure and function of stalk-strut and C sequence. **a**, Knobs-into-holes side-chain packing in stalk-strut coiled coils of Δ MTBD-A and Δ MTBD-B. Side chains of residues, assigned as knobs, are shown as red space-filled models. **b**, Open-book presentation of stalk-strut interactions in Δ MTBD-A. Residues involved in interactions are shown. Truncation positions (Q3824 and E3864) in Δ SD are indicated. **c**, Microtubule-activated ATPase rates. Each symbol indicates mean \pm s.d. ($n = 3$). The grey dotted lines indicate wild-type rates. **d**, Microtubule-binding activity with or without ATP. The smooth curves are the best fits to hyperbolas; wild type (dotted lines) exhibited weak ($K_d > 10 \mu\text{M}$) and strong ($K_d \sim 0.4 \mu\text{M}$) binding to microtubules in the presence and absence of ATP, respectively, whereas Δ SD and Δ C-seq showed weak MT binding ($K_d > 10 \mu\text{M}$) regardless of the presence of ATP. **e**, Back side view of the motor domain, showing the C-sequence H1 helix that interacts with the H6 helix of AAA5 connected to the strut.

Fig. 8) may directly affect the position and/or orientation of H2 and PS-I inserts protruding from AAA2 and thereby may drive the linker swing.

The allosteric communication pathway

Another key question regarding the dynein mechanism is how the motor coordinates microtubule binding at MTBD with the primary ATPase at AAA1, which must be communicated over a long distance through several structural units including the stalk coiled coil. In the present structures, the stalk-strut exhibits the most significant structural variation between the four molecules. Δ MTBD-A has a straight stalk made of a canonical antiparallel coiled coil along its entire length and a strut without canonical coiled-coil packing (left, Fig. 4a and Supplementary Fig. 11). In contrast, Δ MTBD-B possesses a kinked stalk with a shorter coiled-coil region and a strut with canonical coiled-coil packing (right, Fig. 4a and Supplementary Fig. 11). In

WT-A, the distal portion of the inward helix of the stalk appears to melt just below the junction to MTBD, in contrast to the fully folded stalk in Δ MTBD (Supplementary Fig. 12). These structural differences suggest that the stalk–strut structure has a high degree of molecular plasticity that may be relevant to its function of relaying two-way allosteric communication between MTBD and the ring: the structural information could be propagated through the local melting and/or tilting of the stalk, in addition to the recently proposed ‘helix-sliding mechanism’^{30–32}.

The distal end of the strut (residues 3831–3862) interacts with the centre of the stalk (residues 3284–3302 and 3556–3574) to form a four-helix bundle that is stabilized by three hydrogen bonds and hydrophobic interactions involving 29 residues with a wide interaction interface of approximately 800 Å² (Fig. 4b). The extensive interactions suggest that the strut has a key role in the stalk-mediated allosteric communication. To examine this, we performed functional analysis of a strut mutant (Δ SD) generated by replacing its distal portion (Q3824–E3864) with a flexible linker (GGGG). With this mutation, the interaction interfaces of the strut with the stalk were completely removed (Fig. 4b). This mutant showed microtubule-independent, high ATPase activity that is relatively close to the k_{cat} value of wild-type microtubule-activated ATPase activity ($\sim 90 \text{ s}^{-1}$; Fig. 4c) and exhibited an ATP-insensitive, weak microtubule-binding affinity (dissociation constant (K_d) $> 10 \mu\text{M}$; Fig. 4d). The results indicate that the communication between MTBD and the AAA1 ATPase site was disrupted by the mutation, and highlights the importance of stalk–strut interactions for allosteric communication.

We next examined how structural changes propagate from the stalk–strut to the primary AAA1 ATPase site. We previously proposed involvement of the C sequence¹⁷. To test this and to identify the functionally crucial C-sequence segment, we designed a deletion mutant (Δ C-seq) lacking most of the C sequence (S4416–I4730) based on the present high-resolution structure (Fig. 4e and Supplementary Fig. 13). This mutant showed microtubule-independent, low ATPase activity close to the wild-type basal rate and ATP-insensitive, weak microtubule-binding affinity (Fig. 4c, d and Supplementary Fig. 10b), illustrating the importance of the C sequence for allosteric communication. Together with our previous finding that a C-terminal segment

of the C sequence (S4448–I4730) is dispensable for dynein’s motor activity³³, the results indicate that the remaining H1 helix (E4412–D4439) is crucial for communication (Fig. 4e and Supplementary Fig. 13). Indeed, the H1 helix forms a structural bridge between the AAA1- α/β –AAA6- α segment and the basal portion of the strut (Fig. 4e and Supplementary Fig. 13): this helix extends from AAA6- α adjacent to AAA1- α/β , forms a sandwiched structure with AAA5- α and AAA5 extension, and makes direct contacts with the H6 helix of AAA5 from which the strut extends towards the stalk. Collectively, our structure and functional analyses suggest that the stalk, strut and C sequence are involved in the allosteric communication pathway between MTBD and the AAA1 ATPase site.

Concluding remarks

The dynein motor domain structure at 2.8 Å resolution reveals the unique composite architecture of the AAA⁺ ATPase ring and dynein-specific force-generating units (linker, stalk–strut and C-sequence). It also provides new insights into how dynein harnesses the unique architecture to generate force and movement along microtubules. We propose that dynein uses the two distinct molecular machineries embedded on the different ring faces to execute the two fundamental motor actions required for dynein motility: driving of the linker swing for the power stroke and allosteric regulation of MTBD for modulation of microtubule-binding affinity (Fig. 5). On the ring’s front face, the linker swing could be driven by similar mechanisms as those of remodelling actions in other AAA⁺ ATPases through the ATPase-induced open–closed transition of the interfaces between the adjacent AAA1–AAA2 modules and the critical participation of the remodelling apparatus, H2 and PS-I inserts of AAA2. On the ring’s back face, the long-range allosteric communication between AAA1 and MTBD required for coupling of the ATPase cycle with the switching of microtubule affinity could be relayed through dynein-specific structures, that is, stalk, strut and C sequence. The general AAA⁺ ATPase-type and dynein-specific actions coordinated by the primary ATPase cycle at AAA1 would drive dynein motility.

METHODS SUMMARY

For crystallization, we used HF380 in which a His₆–Flag tag was attached to the 5’ terminus of the 380-kDa gene. To create Δ MTBD, the MTBD coding region (A3372–K3495) of HF380 was replaced by a linker sequence (encoding TG). To produce Δ H2I and Δ SD, H2I (S2319–D2329) and strut (Q3824–E3864) coding regions were replaced by a linker sequence (encoding GG and GGGG for Δ H2I and Δ SD, respectively). To construct Δ C-seq, the 3’-terminal region encoding S4416–I4730 was removed. Expression and purification of the motor domains were performed as described^{17,19}.

Crystallization of the motor domain was performed essentially as described¹⁷. All X-ray diffraction data were collected at 100 K on the beamline BL44XU of SPring-8, Harima, Japan. Phase calculations proceeded successfully with the autoSHARP pipeline³⁴. An atomic model of Δ MTBD was built on the density-modified electron-density map and refined at 2.8 Å resolution. The wild-type structure was determined using the molecular replacement method with the model of Δ MTBD. MTBD of chain A was clearly seen in the electron-density map, and a homology model of MTBD was located in the density. The atomic coordinates of wild type were further refined by CNS³⁵ at 3.8 Å resolution with the assistance of deformable elastic network (DEN) restraints³⁶.

ATPase rates²⁰, microtubule binding³⁷ and FRET¹⁵ were measured as described. Phosphate release was measured using a PBP assay³⁸ with an SX-18MV stopped-flow apparatus (Applied Photophysics).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 November 2011; accepted 17 February 2012.

Published online 7 March 2012.

1. Höök, P. & Vallee, R. B. The dynein family at a glance. *J. Cell Sci.* **119**, 4369–4371 (2006).
2. Karki, S. & Holzbaur, E. L. Cytoplasmic dynein and dynactin in cell division and intracellular transport. *Curr. Opin. Cell Biol.* **11**, 45–53 (1999).

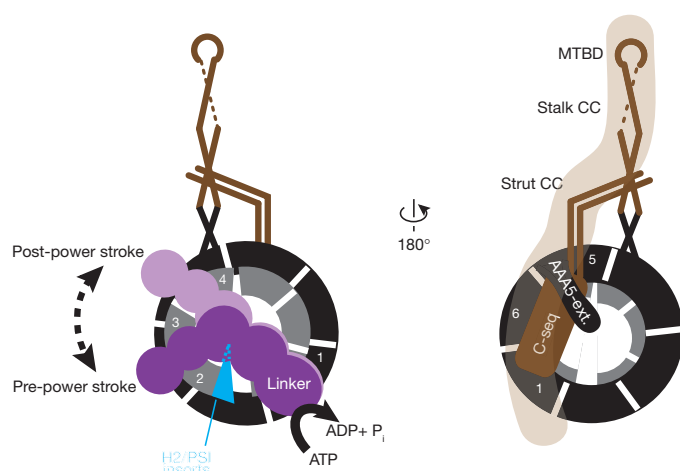


Figure 5 | Proposed mechanism underlying dynein’s motor actions. On the front face of the ring (left), the linker swing for dynein’s power stroke is driven by a mechanism similar to remodelling actions observed in other AAA⁺ ATPases. The primary AAA1 ATPase cycle-driven open–closed transitions of the AAA1–AAA2 interfaces affect the position and/or orientation of H2 and PS-I inserts of AAA2; the structural changes then induce linker bending motions, possibly around the subdomains 2/3 junction, leading to the linker swinging against the ring. On the ring’s back face (right), the long-range two-way allosteric communication between AAA1 and MTBD is relayed through dynein-specific structural units (C-sequence, strut and stalk) to couple the primary ATPase cycle with regulation of microtubule-binding affinity.

3. Vallee, R. B., Williams, J. C., Varma, D. & Barnhart, L. E. Dynein: An ancient motor protein involved in multiple modes of transport. *J. Neurobiol.* **58**, 189–200 (2004).
4. Scholey, J. M. Intraflagellar transport motors in cilia: moving along the cell's antenna. *J. Cell Biol.* **180**, 23–29 (2008).
5. Gibbons, I. R. Cilia and flagella of eukaryotes. *J. Cell Biol.* **91**, 107–124 (1981).
6. DiBella, L. M. & King, S. M. Dynein motors of the *Chlamydomonas* flagellum. *Int. Rev. Cytol.* **210**, 227–268 (2001).
7. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA⁺: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
8. Tucker, P. A. & Sallai, L. The AAA⁺ superfamily—a myriad of motions. *Curr. Opin. Struct. Biol.* **17**, 641–652 (2007).
9. Hanson, P. I. & Whiteheart, S. W. AAA⁺ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
10. Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J. & Oiwa, K. Dynein structure and power stroke. *Nature* **421**, 715–718 (2003).
11. Samsó, M. & Koonce, M. P. 25 Å resolution structure of a cytoplasmic dynein motor reveals a seven-member planar ring. *J. Mol. Biol.* **340**, 1059–1072 (2004).
12. Roberts, A. J. *et al.* AAA⁺ ring and linker swing mechanism in the dynein motor. *Cell* **136**, 485–495 (2009).
13. Gee, M. A., Heuser, J. E. & Vallee, R. B. An extended microtubule-binding structure within the dynein motor domain. *Nature* **390**, 636–639 (1997).
14. Koonce, M. P. Identification of a microtubule-binding domain in a cytoplasmic dynein heavy chain. *J. Biol. Chem.* **272**, 19714–19718 (1997).
15. Kon, T., Mogami, T., Ohkura, R., Nishiura, M. & Sutoh, K. ATP hydrolysis cycle-dependent tail motions in cytoplasmic dynein. *Nature Struct. Mol. Biol.* **12**, 513–519 (2005).
16. Carter, A. P., Cho, C., Jin, L. & Vale, R. D. Crystal structure of the dynein motor domain. *Science* **331**, 1159–1165 (2011).
17. Kon, T., Sutoh, K. & Kurisu, G. X-ray structure of a functional full-length dynein motor domain. *Nature Struct. Mol. Biol.* **18**, 638–642 (2011).
18. Koonce, M. P. & Samsó, M. Overexpression of cytoplasmic dynein's globular head causes a collapse of the interphase microtubule network in *Dictyostelium*. *Mol. Biol. Cell* **7**, 935–948 (1996).
19. Kon, T., Shima, T. & Sutoh, K. Protein engineering approaches to study the dynein mechanism using a dictyostelium expression system. *Methods Cell Biol.* **92**, 65–82 (2009).
20. Kon, T., Nishiura, M., Ohkura, R., Toyoshima, Y. Y. & Sutoh, K. Distinct functions of nucleotide-binding/hydrolysis sites in the four AAA modules of cytoplasmic dynein. *Biochemistry* **43**, 11266–11274 (2004).
21. Iyer, L. M., Leippe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA⁺ ATPases. *J. Struct. Biol.* **146**, 11–31 (2004).
22. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA⁺ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
23. Mocz, G. & Gibbons, I. R. Phase partition analysis of nucleotide binding to axonemal dynein. *Biochemistry* **35**, 9204–9211 (1996).
24. Paschal, B. M. & Vallee, R. B. Retrograde transport by the microtubule-associated protein MAP 1C. *Nature* **330**, 181–183 (1987).
25. Shpetner, H. S., Paschal, B. M. & Vallee, R. B. Characterization of the microtubule-activated ATPase of brain cytoplasmic dynein (MAP 1C). *J. Cell Biol.* **107**, 1001–1009 (1988).
26. Ogura, T., Whiteheart, S. W. & Wilkinson, A. J. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA⁺ ATPases. *J. Struct. Biol.* **146**, 106–112 (2004).
27. Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).
28. Suno, R. *et al.* Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* **22**, 575–585 (2006).
29. Enemark, E. J. & Joshua-Tor, L. On helicases and other motor proteins. *Curr. Opin. Struct. Biol.* **18**, 243–257 (2008).
30. Gibbons, I. R. *et al.* The affinity of the dynein microtubule-binding domain is modulated by the conformation of its coiled-coil stalk. *J. Biol. Chem.* **280**, 23960–23965 (2005).
31. Carter, A. P. *et al.* Structure and functional role of dynein's microtubule-binding domain. *Science* **322**, 1691–1695 (2008).
32. Kon, T. *et al.* Helix sliding in the stalk coiled coil of dynein couples ATPase and microtubule binding. *Nature Struct. Mol. Biol.* **16**, 325–333 (2009).
33. Numata, N., Shima, T., Ohkura, R., Kon, T. & Sutoh, K. C-sequence of the *Dictyostelium* cytoplasmic dynein participates in processivity modulation. *FEBS Lett.* **585**, 1185–1190 (2011).
34. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
35. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
36. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
37. Imamula, K., Kon, T., Ohkura, R. & Sutoh, K. The coordination of cyclic microtubule association/dissociation and tail swing of cytoplasmic dynein. *Proc. Natl Acad. Sci. USA* **104**, 16134–16139 (2007).
38. White, H. D., Belknap, B. & Webb, M. R. Kinetics of nucleoside triphosphate cleavage and phosphate release steps by associated rabbit skeletal actomyosin, measured using a novel fluorescent probe for phosphate. *Biochemistry* **36**, 11828–11836 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Yamashita, Y. Umena, M. Suzuki and A. Nakagawa of SPring-8 BL-44XU for their support during X-ray data collection; and T. Kikuchi and R. Ohkura for their technical support. We are grateful to C. Toyoshima for discussion of X-ray data collection; K. Kinosita Jr and T. Tsukihara for their support and encouragement. This work was supported by Grants-in-Aid for Scientific Research (17770126, 20687011 and 23370073 (T.K.), 16083205 and 17107003 (K.S.), 17053006, 18054008 and 20051006 (G.K.)) from the Ministry of Education, Culture Sports, Science, and Technology of Japan and a grant from the Human Frontier Science Program (T.K.).

Author Contributions T.K., K.S. and G.K. designed the study. T.K. purified, crystallized and collected X-ray data; T.O. and G.K. processed and refined X-ray data; T.K., R.S.-K., K.I. and T.S. performed functional analyses; T.K., K.S. and G.K. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 3VKG (ΔMTBD) and 3VKH (wild type). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.K. (takahide.kon@protein.osaka-u.ac.jp).

METHODS

Protein engineering, expression and purification. All constructs were generated using standard polymerase chain reaction (PCR)-based cloning methods. We used the 380-kDa motor domain (V1388–I4730; the numbering is based on DDB0185096 in dictyBase and is different from the numbering we used previously)^{18,19} of the cytoplasmic dynein heavy chain isolated from *D. discoideum* Ax2 strain. For crystallization, we used the HF380 construct in which a His₆-Flag tag (encoding MTRHHHHHHGGGDKDDDDKGGGK) was attached to the 5' terminus of the 380-kDa gene. We used the HFB380 construct³⁷ for phosphate release measurements and the HFG380B2 construct¹⁵ for ATPase, microtubule-binding and FRET measurements. To create the Δ MTBD mutant, the MTBD coding region (A3372–K3495) of HF380 was replaced by a linker sequence (encoding TG). To produce the Δ H2I and Δ SD mutants, the H2I coding region (S2319–D2329) and the strut coding region (Q3824–E3864) of HFG380B2 were replaced by a linker sequence (encoding GG and GGGG for Δ H2I and Δ SD, respectively). The previously used Δ C mutant¹² (V1388–I4324) has been found to have the C-terminal truncation beyond the boundary between AAA6 and the C-sequence; we therefore re-designed the Δ C-seq mutant (V1388–E4415) by removing the 3'-terminal region of the 380-kDa gene encoding S4416–I4730 from HFG380B2. Each of the constructs was subcloned into the MB38 plasmid³⁹. After verification of the nucleotide sequence of the PCR-amplified regions, each of the expression constructs was introduced into *D. discoideum* cells; subsequently, the transformed cells were selected, cultivated and collected as described¹⁹. For the production of the L-selenomethionine-containing motor domain, the *D. discoideum* cells were cultivated in HL5 medium (ForMedium) supplemented with 10 μ g ml⁻¹ blasticidin S, 10 μ g ml⁻¹ G418 and 12 μ g ml⁻¹ tetracycline at 22 °C with shaking at 150 r.p.m. until the cell density reached $\sim 1 \times 10^7$ cells per ml; collected by centrifugation at 1,320g for 5 min; re-suspended in SIH medium without methionine (ForMedium) supplemented with 100 μ g ml⁻¹ L-selenomethionine; and then cultivated further at 22 °C with shaking at 150 r.p.m. for ~ 24 h. The expressed motor domain was purified as described^{17,19}. For crystallization, the protein was concentrated to 5.25 mg ml⁻¹ and supplemented with 3 mM ADP. 7-Diethylamino-3-(((2-maleimidyl)ethyl)amino)carbonyl-coumarin-labelled phosphate-binding protein (MDCC-PBP) was prepared as described³⁸. Protein concentrations were determined by the Bradford method standardized with bovine serum albumin. The microtubule concentration is expressed as the tubulin dimer concentration.

Crystallization and structure determination. Crystallization of the motor domain was performed as described¹⁷. For cryo-protection, the crystals were transferred into a start solution (5 mM Bis-tris propane (pH 7.0), 7.5 mM Tris (pH 8.0), 2 mM MgCl₂, 8% PEG 8000, 1.5 mM ADP), and the concentration of PEG 400 was gradually increased from 0% to 30% in 5% increments using the soaking method. The crystals were then flash-frozen in a cold nitrogen gas stream at 100 K. For Ta₆Br₁₂ and PW₁₂O₄₀ derivatization, the crystals were cryo-protected and then soaked in the cryo-protection solution containing 0.4 mM Ta₆Br₁₂ and 0.03 mM PW₁₂O₄₀, respectively, for 3 days at 277 K before freezing.

All X-ray diffraction data were collected at 100 K on the beamline BL44XU of SPring-8, Harima, Japan. Native data were measured at the wavelength of 0.9000 Å, and the Ta₆Br₁₂, PW₁₂O₄₀ and L-selenomethionine-containing heavy atom derivative data were measured at the wavelength of 1.2547, 1.2142 and 0.9790 Å, respectively. X-ray diffraction images were processed using the HKL2000 program package⁴⁰. The structure was determined by the multiple isomorphous replacement method with anomalous scattering (MIRAS) using the three heavy atom derivative data. The refinement of heavy-atom positions was performed with SHARP⁴¹. Phase calculations proceeded successfully with the autoSHARP pipeline³⁴, with the reasonable figure of merit values before density modification (0.18 for centric and 0.26 for acentric reflections). An initial atomic model of Δ MTBD was built on the density-modified electron-density map using COOT⁴². An electron-density map was further improved by using non-crystallographic symmetry (NCS) averaging. The crystallographic refinement was performed by CNS³⁵ at 2.8 Å resolution, and then the manual model correction and refinement were iterated until convergence. The wild-type structure was determined using the molecular replacement method with the atomic model of Δ MTBD as a probe. Several rounds of rigid-body refinement gave a crystallographic

R_{work} and R_{free} below 40%. At this stage, the MTBD of chain A was clearly seen in the electron-density map, and a homology model of the MTBD generated by SWISS-MODEL⁴³ was located in the density. All atomic coordinates were further refined by CNS at 3.8 Å resolution with the assistance of deformable elastic network (DEN) restraints³⁶. The models of Δ MTBD and wild type have no Ramachandran outlier. Intra-molecular interactions were analysed using PISA⁴⁴ and HBPLUS⁴⁵ with LigPlot+⁴⁶. The secondary structure elements were assigned using the program DSSP⁴⁷. We identified and analysed the coiled-coil interactions with SOCKET⁴⁸, and also checked them with TWISTER⁴⁹. All figures showing the atomic coordinates were made with PYMOL⁵⁰.

Steady-state biochemical assays. ATPase rates of HFG380B2 and its mutants (2–30 nM) were measured in a dynein assay buffer (10 mM K-PIPES, 50 mM potassium acetate, 4 mM MgSO₄, 1 mM EGTA, and 1 mM DTT, pH 7.0) supplemented with 10 μ M paclitaxel and the indicated concentration of microtubules using a coupled enzymatic assay (EnzChek phosphate assay kit, Invitrogen) at 25 °C after mixing with 1 mM ATP, as described²⁰. Note that we used the monomeric motor domain constructs for the functional analyses, and thus the K_m (microtubule) value for microtubule-activated ATPase (~ 30 μ M for wild-type constructs²⁰) is higher than those reported for dimeric constructs³³. Microtubule binding was measured by co-sedimentation with microtubules in the absence or presence of 3.3 mM ATP in the assay buffer supplemented with 5 μ M free paclitaxel and 250 μ M GTP at 25 °C, as described³⁷. To determine FRET efficiencies, fluorescence intensities of the BFP and GFP moieties of HFG380B2 or its mutant (~ 120 nM) were measured in the assay buffer at 25 °C after mixing with 200 μ M of the indicated nucleotide, and then FRET efficiencies were calculated from the sensitized emission of the acceptor (GFP) as described¹⁵.

Pre-steady-state phosphate release measurements. Pre-steady-state kinetic measurements were performed at 25 °C using a PBP assay³⁸ with an SX-18MV stopped-flow apparatus (Applied Photophysics). For the measurements, the motor domain (1 μ M) in the assay buffer supplemented with 0.1 U ml⁻¹ purine nucleoside phosphorylase, 500 μ M 7-methylguanosine, and 5 μ M MDCC-PBP, was mixed with the same volume of a substrate solution (2 mM ATP and 40 μ M microtubules) in the assay buffer supplemented with 0.1 U ml⁻¹ purine nucleoside phosphorylase, 500 μ M 7-methylguanosine, 5 μ M MDCC-PBP, and 10 μ M free paclitaxel. Phosphate release from the motor domain was followed by exciting the reaction mixture at 425 nm and monitoring the MDCC-PBP fluorescence enhancement using a 455-nm cutoff filter.

39. Blaauw, M., Linskens, M. H. & van Haastert, P. J. Efficient control of gene expression by a tetracycline-dependent transactivator in single *Dictyostelium discoideum* cells. *Gene* **252**, 71–82 (2000).
40. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
43. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
44. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
45. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
46. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
47. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
48. Walshaw, J. & Woolfson, D. N. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.* **307**, 1427–1450 (2001).
49. Strelkov, S. V. & Burkhard, P. Analysis of α -helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **137**, 54–64 (2002).
50. Schrödinger, L. L. C., *The PyMOL Molecular Graphics System, Version 1.3r1* (2010).

An absence of neutrinos associated with cosmic-ray acceleration in γ -ray bursts

IceCube Collaboration*

Very energetic astrophysical events are required to accelerate cosmic rays to above 10^{18} electronvolts. GRBs (γ -ray bursts) have been proposed as possible candidate sources^{1–3}. In the GRB ‘fireball’ model, cosmic-ray acceleration should be accompanied by neutrinos produced in the decay of charged pions created in interactions between the high-energy cosmic-ray protons and γ -rays⁴. Previous searches for such neutrinos found none, but the constraints were weak because the sensitivity was at best approximately equal to the predicted flux^{5–7}. Here we report an upper limit on the flux of energetic neutrinos associated with GRBs that is at least a factor of 3.7 below the predictions^{4,8–10}. This implies either that GRBs are not the only sources of cosmic rays with energies exceeding 10^{18} electronvolts or that the efficiency of neutrino production is much lower than has been predicted.

Neutrinos from GRBs are produced in the decay of charged pions produced in interactions between high-energy protons and the intense γ -ray background within the GRB fireball, for example in the Δ -resonance process $p + \gamma \rightarrow \Delta^+ \rightarrow n + \pi^+$ (p , proton; γ , photon (here γ -ray); Δ^+ , delta baryon; n , neutron; π^+ , pion). When these pions decay via $\pi^+ \rightarrow \mu^+ \nu_\mu$ and $\mu^+ \rightarrow e^+ \nu_e \bar{\nu}_\mu$, they produce a flux of high-energy muon neutrinos (ν_μ) and electron neutrinos (ν_e), coincident with the γ -rays, and peaking at energies of several hundred tera-electronvolts (TeV)^{4,11} (μ^+ , antimuon; e^+ , positron). Such a flux should be detectable using km³-scale instruments like the IceCube neutrino telescope^{12,13} (Supplementary Fig. 1). Owing to maximal mixing between muon neutrinos and tau neutrinos, neutrinos from pion decay in and around GRBs will arrive at Earth in an equal mixture of flavours. We focus here only on muons produced in ν_μ charged-current interactions. As the downgoing cosmic-ray muon background presents challenges for the identification of neutrino-induced muons, we achieve our highest sensitivity for upgoing neutrinos (from sources in the northern sky). However, the tight constraint of spatial and temporal coincidence with a GRB allows some sensitivity even in the southern sky. One of the two analyses presented here therefore includes Southern Hemisphere GRBs during the 2009–10 IceCube run.

The results presented here were obtained while IceCube was under construction, using 40 and 59 of the 86 photomultiplier strings of the final detector (Supplementary Fig. 1), which took data from April 2008 to May 2009 and from May 2009 until May 2010, respectively. During the 59-string data-taking period, 190 GRBs were observed and reported by γ -ray observatory satellites via the GRB Coordinates Network¹⁴, with 105 in the northern sky. Of those GRBs, 9 were not included in our catalogue owing to detector downtime associated with construction and calibration. Two additional GRBs were included from test runs before the start of the official 59-string run. 117 northern-sky GRBs were included from the 40-string period⁷ to compute the final combined result. GRB positions were taken from the satellite with the smallest reported error, which is typically smaller than the IceCube resolution. The GRB γ -ray emission start (T_{start}) and stop (T_{stop}) times were taken by finding the earliest and latest time reported for γ -ray emission.

As in our previous study⁷, we conducted two analyses of the IceCube data. In a model-dependent search, we examine data during the period of γ -ray emission reported by any satellite for neutrinos with the energy spectrum predicted from the γ -ray spectra of individual GRBs^{6,9}. The model-independent analysis searches more generically for neutrinos on wider timescales, up to the limit of sensitivity to small numbers of events at ± 1 day, or with different spectra. Both analyses follow the methods used in our previous work⁷, with the exception of slightly changed event selection and the addition of the Southern Hemisphere to the model-independent search. Owing to the large background of downgoing muons from the southern sky, the Southern Hemisphere analysis is sensitive mainly to higher-energy events (Supplementary Fig. 3). Systematic uncertainties from detector effects have been included in the reported limits from both analyses, and were estimated by varying the simulated detector response and recomputing the limit, with the dominant factor being the efficiency of the detector’s optical sensors.

In the 59-string portion of the model-dependent analysis, no events were found to be both on-source and on time (within 10° of a GRB and between T_{start} and T_{stop}). From the individual burst spectra^{6,9} with an assumed ratio of energy in protons to energy in electrons $\varepsilon_p/\varepsilon_e = 10$ (ref. 6), 8.4 signal events were predicted from the combined 2-year data set and a final upper limit (90% confidence) of 0.27 times the predicted flux can be set (Fig. 1). This corresponds to a 90% upper limit on $\varepsilon_p/\varepsilon_e$ of 2.7, with other parameters held fixed, and includes a 6% systematic uncertainty from detector effects.

In the model-independent analysis, two candidate events were observed at low significance, one 30 s after GRB 091026A (event 1) and another 14 h before GRB 091230A (most theories predict neutrinos within a few minutes of the burst). Subsequent examination showed they had both triggered several tanks in the IceTop surface air shower array, and are thus very probably muons from cosmic-ray air showers. In Fig. 2 are shown limits from this analysis on the normalization of generic power-law muon neutrino spectra expected from shock acceleration at Earth as a function of the size of the time window $|\Delta t|$, which is the difference between the neutrino arrival time and the first reported satellite trigger time. As a cross-check on both results, the limit from this analysis on the average individual burst spectra^{6,9} during the time window corresponding to the median duration of the bursts in the sample (28 s) was 0.24 times the predicted flux, within 10% of the model-dependent analysis.

Assuming that the GRBs in our catalogue are a representative sample of a total of 667 per year (ref. 7), we can scale the emission from our catalogue to the emission of all GRBs. The resulting limits can then be compared to the expected neutrino rates from models that assume that GRBs are the main sources of ultra-high-energy cosmic rays^{4,8,10}, with sampling biases of the same order as model uncertainties in the flux predictions^{15,16}. Limits from the model-independent analysis on fluxes of this type are shown in Fig. 3.

These limits exclude all tested models^{4,8–10} with their standard parameters and uncertainties on those parameters (Figs 1, 3). The models are different formulations of the same fireball phenomenology,

*Lists of participants and their affiliations appear at the end of the paper.

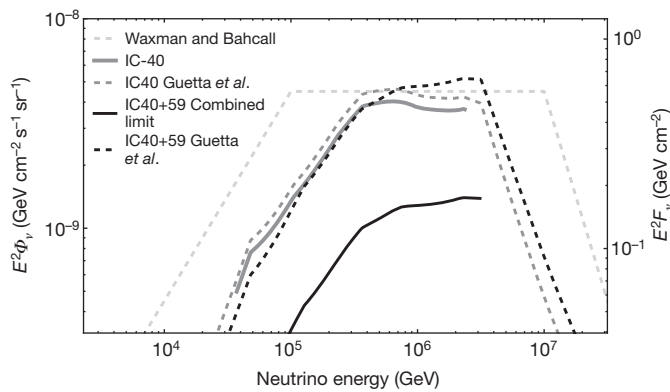


Figure 1 | Comparison of results to predictions based on observed γ -ray spectra. The summed flux predictions normalized to γ -ray spectra^{6,9,19} are shown as a function of neutrino energy (E) in dashed lines, with the dark grey dashed line labelled 'IC40 Guetta *et al.*' showing the flux prediction for the 40-string portion of the analysis, and the black dashed line labelled 'IC40+59 Guetta *et al.*' showing the prediction for the full two-year dataset. The cosmic ray normalized Waxman-Bahcall flux^{4,20} is also shown for reference as the pale grey dashed line. 90% confidence upper limits on these spectra are shown as solid lines, with the grey line labelled 'IC40 limit' showing the previous IceCube result⁶ and the black 'IC40+IC59 Combined' line showing the result from the full dataset (this work). The predicted neutrino flux, when normalized to the γ -rays^{6,9}, is proportional to the ratio of energy in protons to that in electrons, which are presumed responsible for the γ -ray emission (ϵ_p/ϵ_e , here the standard 10). The flux shown is slightly modified⁶ from the original calculation⁹. Φ_v (left vertical axis) is the average neutrino flux at Earth, obtained by scaling the summed predictions from the bursts in our sample (F_v , right vertical axis) by the global GRB rate (here 667 bursts yr^{-1} ; ref. 7). The first break in the neutrino spectrum is related to the break in the photon spectrum measured by the satellites, and the threshold for photo-pion production, whereas the second break corresponds to the onset of synchrotron losses of muons and pions. Not all of the parameters used in the neutrino spectrum calculation are measurable from every burst. In such cases, benchmark values⁷ were used for the unmeasured parameters. Data shown here were taken from the result of the model-dependent analysis.

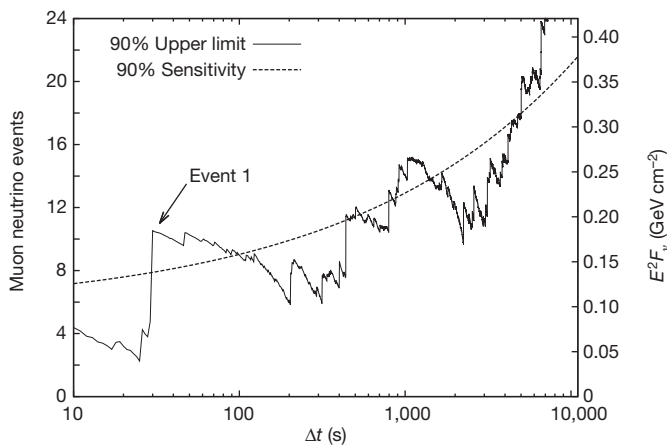


Figure 2 | Upper limits on E^{-2} power-law muon neutrino fluxes. Limits were calculated using the Feldman-Cousins method²¹ from the results of the model-independent analysis. The left-hand y-axis shows the total number of expected ν_μ events, while the right-hand y-axis (F_v) is the same as in Fig. 1. A time window of Δt implies observed events arriving between t seconds before the burst and t afterward. The variation of the upper limit (solid line labelled '90% Upper limit') with Δt reflects statistical fluctuations in the observed background rate, as well as the presence of individual events of varying quality. The dashed line labelled '90% Sensitivity' shows the upper limit that would have been obtained with exactly the mean expected background. The event at 30 s (event 1) is consistent with background and believed to be a cosmic-ray air shower.

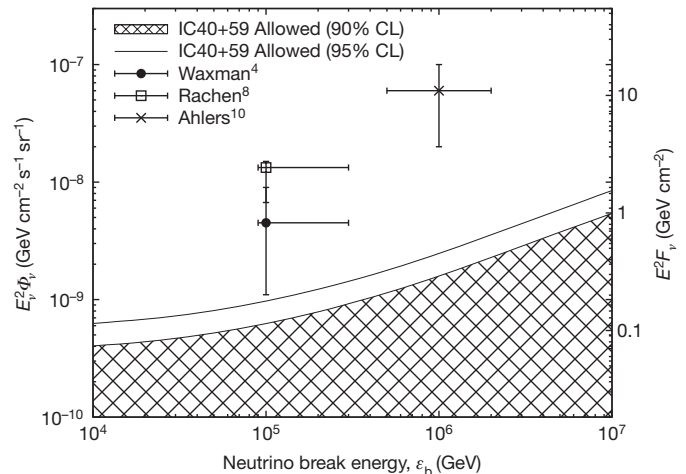


Figure 3 | Compatibility of some neutrino flux predictions based on cosmic ray production in GRBs with observations. The cross-hatched area ('IC40+59 Allowed 90% CL') shows the 90% confidence allowed values of the neutrino flux (vertical axes, as in Fig. 1) versus the neutrino break energy (ϵ_b) in comparison to model predictions with estimated uncertainties (points); the solid line labelled 'IC40+59 Allowed 95% CL' shows the upper bound of the 95% confidence allowed region. Data were taken from the model-independent analysis from the time window corresponding to the median duration of the GRBs in our catalogue ($|\Delta t| = 28$ s). Spectra are represented here as broken power laws ($\Phi_v \cdot \{E^{-1}/\epsilon_b, E < \epsilon_b; E^{-2}, E > \epsilon_b\}$) with a break energy ϵ_b corresponding to the Δ resonance for $p-\gamma$ interactions in the frame of the shock. The muon flux in IceCube is dominated by neutrinos with energies around the first break (ϵ_b). As such, the upper break, due to synchrotron losses of π^+ , has been neglected here, as its presence or absence does not contribute significantly to the muon flux and thus does not have a significant effect on the presented limits. ϵ_b is related to the bulk Lorentz factor Γ ($\epsilon_b \propto \Gamma^2$); all of the models shown assume $\Gamma \approx 300$. The value of Γ corresponding to 10^7 GeV is $>1,000$ for all models. Vertical axes are related to the accelerated proton flux by the model-dependent constant of proportionality f_π . For models assuming a neutron-decay origin of cosmic rays (ref. 8 and ref. 10) f_π is independent of Γ ; for others (ref. 4) $f_\pi \propto \Gamma^{-4}$. Error bars on model predictions are approximate and were taken either from the original papers, where included¹⁰, or from the best-available source in the literature¹⁵ otherwise. The errors are due to uncertainties in f_π and in fits to the cosmic-ray spectrum. Waxman-Bahcall⁴ (circle) and Rachen⁸ (box) fluxes were calculated using a cosmic-ray density of $(1.5-3) \times 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$, with 3×10^{44} the central value²⁰. The Ahlers¹⁰ model is shown with a cross. CL, confidence level.

producing neutrinos at proton-photon ($p-\gamma$) interactions in internal shocks. The remaining parameter spaces available to each model therefore have similar characteristics: either a low density of high-energy protons, below that required to explain the cosmic rays, or a low efficiency of neutrino production.

In the GRB fireball, protons are believed to be accelerated stochastically in collisions of internal shocks in the expanding GRB. The neutrino flux is proportional to the rate of $p-\gamma$ interactions, and so to the proton content of the burst by a model-dependent factor. Assuming a model-dependent proton ejection efficiency, the proton content can in turn be related to the measured flux of high-energy cosmic rays if GRBs are the cosmic-ray sources. Limits on the neutrino flux for cosmic-ray-normalized models are shown in Fig. 3; each model prediction has been normalized to a value consistent with the observed ultra-high-energy cosmic-ray flux. The proton density can also be expressed as a fraction of the observed burst energy, directly limiting the average proton content of the bursts in our catalogue (Fig. 4).

An alternative is to reduce the neutrino production efficiency, for example by modifying the physics included in the predictions^{16,17} or by increasing the bulk Lorentz boost factor, Γ . Increasing Γ increases the proton energy threshold for pion production in the observer frame, thereby reducing the neutrino flux owing to the lower proton density at higher energies. Astrophysical lower limits on Γ are established by pair

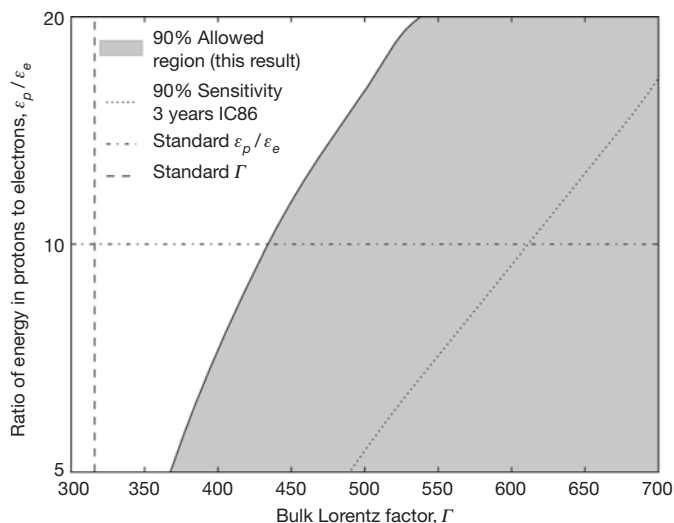


Figure 4 | Constraints on fireball parameters. The shaded region, based on the result of the model dependent analysis, shows the values of GRB energy in protons and the average fireball bulk Lorentz factor for modelled fireballs^{6,9} allowed by this result at the 90% confidence level. The dotted line indicates the values of the parameters to which the completed IceCube detector (IC86) is expected to be sensitive after 3 years of data. The standard values considered⁹ are shown as dashed-dotted and dashed lines and are excluded by this analysis. Note that the quantities shown here are model-dependent.

production arguments⁹, but the upper limit is less clear. Although it is possible that Γ may take values of up to 1,000 in some unusual bursts, the average value is probably lower (usually assumed^{6,9} to be around 300) and the non-thermal γ -ray spectra from the bursts set a weak constraint that $\Gamma \lesssim 2,000$ (ref. 18). For all considered models, with uniform fixed proton content, very high average values of Γ are required to be compatible with our limits (Figs 3, 4).

In the case of models where cosmic rays escape from the GRB fireball as neutrons^{8,10}, the neutrons and neutrinos are created in the same $p\text{--}\gamma$ interactions, directly relating the cosmic-ray and neutrino fluxes and removing many uncertainties in the flux calculation. In these models, Γ also sets the threshold energy for production of cosmic rays. The requirement that the extragalactic cosmic rays be produced in GRBs therefore does set a strong upper limit on Γ : increasing it beyond $\sim 3,000$ causes the proton flux from GRBs to disagree with the measured cosmic-ray flux above 4×10^{18} eV, where extragalactic cosmic rays are believed to be dominant. Limits on Γ in neutron-origin models from this analysis ($\gtrsim 2000$, Fig. 3) are very close to this point, and as a result all such models—in which all extragalactic cosmic rays are emitted from GRBs as neutrons—are now largely ruled out.

Although the precise constraints are model-dependent, the general conclusion is the same for all the versions of fireball phenomenology we have considered here: either the proton density in GRB fireballs is substantially below the level required to explain the highest-energy cosmic rays or the physics in GRB shocks is significantly different from that included in current models. In either case, our current theories of cosmic-ray and neutrino production in GRBs will need to be revisited.

Received 6 January; accepted 8 March 2012.

- Waxman, E. Cosmological gamma-ray bursts and the highest energy cosmic rays. *Phys. Rev. Lett.* **75**, 386–389 (1995).
- Vietri, M. The acceleration of ultra-high-energy cosmic rays in gamma-ray bursts. *Astrophys. J.* **453**, 883–889 (1995).
- Milgrom, M. & Usov, V. Possible association of ultra-high-energy cosmic-ray events with strong gamma-ray bursts. *Astrophys. J.* **449**, L37 (1995).
- Waxman, E. & Bahcall, J. High energy neutrinos from cosmological gamma-ray burst fireballs. *Phys. Rev. Lett.* **78**, 2292–2295 (1997).
- Avrorin, A. V. *et al.* Search for neutrinos from gamma-ray bursts with the Baikal neutrino telescope NT200. *Astron. Lett.* **37**, 692–698 (2011).
- Abbasi, R. *et al.* Search for muon neutrinos from gamma-ray bursts with the IceCube neutrino telescope. *Astrophys. J.* **710**, 346–359 (2010).

- Abbasi, R. *et al.* Limits on neutrino emission from gamma-ray bursts with the 40 string IceCube detector. *Phys. Rev. Lett.* **106**, 141101 (2011).
- Rachen, J. P. & Mészáros, P. in *Fourth Huntsville Gamma-Ray Burst Symposium* (eds Meegan, C. A., Preece, R. D. & Koshut, T. M.) 776–780 (American Institute of Physics Conference Proceedings Vol. 428, 1998).
- Guetta, D., Hooper, D., Alvarez-Muñiz, J., Halzen, F. & Reuveni, E. Neutrinos from individual gamma-ray bursts in the BATSE catalog. *Astrophys. J.* **20**, 429–455 (2004).
- Ahlers, M., Gonzalez-Garcia, M. C. & Halzen, F. GRBs on probation: testing the UHE CR paradigm with IceCube. *Astrophys. J.* **35**, 87–94 (2011).
- Becker, J. K. High-energy neutrinos in the context of multimessenger astrophysics. *Phys. Rep.* **458**, 173–246 (2008).
- Abbasi, R. *et al.* The IceCube data acquisition system: signal capture, digitization, and timestamping. *Nucl. Instrum. Methods Phys. Res. A* **601**, 294–316 (2009).
- Ahrens, J. *et al.* Muon track reconstruction and data selection techniques in AMANDA. *Nucl. Instrum. Methods Phys. Res. A* **524**, 169–194 (2004).
- GCN: The Gamma-ray Coordinates Network. <http://gcn.gsfc.nasa.gov>.
- Guetta, D., Spada, M. & Waxman, E. On the neutrino flux from gamma-ray bursts. *Astrophys. J.* **559**, 101–109 (2001).
- Baerwald, P., Hümmer, S. & Winter, W. Systematics in aggregated neutrino fluxes and flavor ratios from gamma-ray bursts. *Astrophys. J.* **35**, 508–529 (2012).
- Hümmer, S., Baerwald, P. & Winter, W. Neutrino emission from gamma-ray burst fireballs, revised. Preprint at <http://arXiv.org/abs/1112.1076> (2011).
- Mészáros, P. Gamma-ray bursts. *Rep. Prog. Phys.* **69**, 2259–2321 (2006).
- Becker, J. K., Stamatiikos, M., Halzen, F. & Rhode, W. Coincident GRB neutrino flux predictions: implications for experimental UHE neutrino physics. *Astrophys. J.* **25**, 118–128 (2006).
- Waxman, E. Astrophysical sources of high energy neutrinos. *Nucl. Phys. B Proc., Suppl.* **118**, 353–362 (2003).
- Feldman, G. J. & Cousins, R. D. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D* **57**, 3873–3889 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from the following agencies: US NSF, Office of Polar Programs, and US NSF, Physics Division; University of Wisconsin Alumni Research Foundation; the GLOW and OSG grids; US DOE, NERSC; the LONI grid; NSERC, Canada; Swedish Research Council, Swedish Polar Research Secretariat, SNIC, K. and A. Wallenberg Foundation, Sweden; German Ministry for Education and Research, Deutsche Forschungsgemeinschaft; Research Department of Plasmas with Complex Interactions (Bochum), Germany; FSR, FWO Odysseus, IWT, BELSPO, Belgium; University of Oxford, UK; Marsden Fund, New Zealand; Australian Research Council; JSPS, Japan; SNSF, Switzerland. J.P.R. was supported by the Capes Foundation, Brazil; N.W. by the NSF GRFP. We thank S. Hümmer, E. Waxman and W. Winter for discussions.

Author Contributions The IceCube observatory was designed and constructed by the IceCube Collaboration and the IceCube Project. It is operated by the IceCube Collaboration, who set science goals. Data processing and calibration, Monte Carlo simulations of the detector and of theoretical models, and data analyses were performed by a large number of IceCube members who also discussed and approved the scientific results. This manuscript was written by P.R. and N.W. and subjected to an internal collaboration-wide review process. All authors approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.W. (nwhitehorn@icecube.wisc.edu) or to P.R. (redlpete@icecube.umd.edu).

IceCube Collaboration R. Abbasi¹, Y. Abdou², T. Abu-Zayyad³, M. Ackermann⁴, J. Adams⁵, J. A. Aguilar⁶, M. Ahlers¹, D. Altmann⁷, K. Andeen¹, J. Auffenberg¹, X. Bai^{8,9}, M. Baker¹, S. W. Barwick¹⁰, R. Bay¹¹, J. L. Bazo Alba⁴, K. Beattie¹², J. J. Beatty^{13,14}, S. Bechet¹⁵, J. K. Becker¹⁶, K.-H. Becker¹⁷, M. Bell¹⁸, M. L. Benabderrahmane⁴, S. BenZvi¹, J. Berdermann⁴, P. Berghaus⁵, D. Berley¹⁹, E. Bernardini⁴, D. Bertrand¹⁹, D. Z. Besson²⁰, D. Bindig¹⁷, M. Bissok⁷, E. Blaufuss¹⁹, J. Blumenthal⁷, D. J. Boersma⁷, C. Böhm²¹, D. Bose²², S. Böser²³, O. Botner²⁴, L. Brayeux²², A. M. Brown⁵, S. Buitink²², K. S. Caballero-Mora¹⁸, M. Carson², M. Casier²², D. Chirkin¹, B. Christy¹⁹, F. Clevermann²⁵, S. Cohen²⁶, C. Colnard²⁷, D. F. Cowen^{18,28}, A. H. Cruz Silva⁴, M. V. D'Agostino¹¹, M. Danninger²¹, J. Daughhetee²⁹, J. C. Davis¹³, C. De Clercq²², T. Degner²³, F. Descamps², P. Desiati¹, G. de Vries-Uiterweerd², T. DeYoung¹⁸, J. C. Díaz-Vélez¹, M. Dierckxsens¹⁵, J. Dreyer¹⁶, J. P. Dumml¹, M. Dunkman¹⁸, J. Eisch¹, R. W. Ellsworth¹⁹, O. Engdegård²⁴, S. Euler⁷, P. A. Evenson⁸, O. Fadiran¹, A. R. Fazely³⁰, A. Fedynitch¹⁶, J. Feintzeig¹, T. Feusels², K. Filimonov¹¹, C. Finley²¹, T. Fischer-Wasels¹⁷, S. Flis²¹, A. Frackowiak²³, R. Franke⁴, T. K. Gaisser³, J. Gallagher³¹, L. Gerhardt^{11,12}, L. Gladstone¹, T. Glusenkamp⁴, A. Goldschmidt¹², J. A. Goodman¹⁹, D. Góra⁴, D. Grant³², T. Griesel³³, A. Groß²⁷, S. Grullon¹, M. Gurtner¹, C. Ha^{11,12}, A. Haj Ismail², A. Hallgren²⁴, F. Halzen¹, K. Han⁴, G. Hanson¹⁵, D. Heereman¹⁵, D. Heinen¹, K. Helbing¹⁷, R. Hellauer¹⁹, S. Hickford⁵, K. C. Hill³⁴, K. D. Hoffman¹⁹, B. Hoffmann⁷, A. Homeier²³, K. Hoshina¹, W. Huelsnitz^{19,35}, J.-P. Hülþ¹, P. O. Hult²¹, K. Hultqvist²¹, S. Hussain⁵, A. Ishihara²⁶, E. Jacobi⁴, J. Jacobsen¹, G. S. Japaridze³⁷, H. Johansson²¹, A. Kappes³⁸, T. Karg¹⁷, A. Karle¹, J. Kiryluk³⁹, F. Kislat⁴, S. R. Klein^{11,12}, J.-H. Köhne²⁵, G. Kohnen⁴⁰, H. Kolanoski³⁸, L. Köpke³³, S. Kopper¹⁷, D. J.

Koskinen¹⁸, M. Kowalski²³, T. Kowarik³³, M. Krasberg¹, G. Kroll³³, J. Kunnen²², N. Kurahashi¹, T. Kuwabara⁸, M. Labare²², K. Laihem⁷, H. Landsman¹, M. J. Larson¹⁸, R. Lauer⁴, J. Lünemann³³, J. Madsen³, A. Marotta¹⁵, R. Maruyama¹, K. Mase³⁶, H. S. Matis¹², K. Meagher¹⁹, M. Merck¹, P. Mészáros^{18,28}, T. Meures¹⁵, S. Miarecki^{11,12}, E. Middell⁴, N. Milke²⁵, J. Miller²⁴, T. Montaruli^{6,41}, R. Morse¹, S. M. Movit²⁸, R. Nahnhauser⁴, J. W. Nam¹⁰, U. Naumann¹⁷, S. C. Nowicki³², D. R. Nygren¹², S. Odrowski²⁷, A. Olivas¹⁹, M. Olivo¹⁶, A. O'Murchadha¹, S. Panknin²³, L. Paul⁷, C. Pérez de los Heros²⁴, A. Piegsa³³, D. Pieloth²⁵, J. Posselt¹⁷, P. B. Price¹¹, G. T. Przybylski¹², K. Rawlins⁴², P. Redl¹⁹, E. Resconi^{27,43}, W. Rhode²⁵, M. Ribordy²⁶, M. Richman¹⁹, B. Riedel¹, A. Rizzo²², J. P. Rodrigues¹, F. Rothmaier³³, C. Rott¹³, T. Ruhe²⁵, D. Rutledge¹⁸, B. Ruzbayev⁸, D. Ryckbosch², H.-G. Sander³³, M. Santander¹, S. Sarkar⁴⁴, K. Schatto³³, T. Schmidt¹⁹, S. Schöneberg¹⁶, A. Schönwald⁴, A. Schukraft⁷, L. Schulte²³, A. Schultes¹⁷, O. Schulz^{27,43}, M. Schunck², D. Seckel⁸, B. Semburg¹⁷, S. H. Seo²¹, Y. Sestayo²⁷, S. Seunarine⁴⁵, A. Silvestri¹⁰, M. W. E. Smith¹⁸, G. M. Spiczak³, C. Spiering⁴, M. Stamatikos^{13,46}, T. Stanev⁸, T. Stezelberger¹², R. G. Stokstad¹², A. Stöbl⁴, E. A. Strahler²², R. Ström²⁴, M. Stürer²³, G. W. Sullivan¹⁹, H. Taavola²⁴, I. Taboada²⁹, A. Tamburro⁸, S. Ter-Antonyan³⁰, S. Tilav⁸, P. A. Toale⁴⁷, S. Toscano¹, D. Tosi⁴, N. van Eijndhoven²², A. Van Overloop², J. van Santen¹, M. Vehring⁷, M. Voge²³, C. Walck²¹, T. Waldenmaier³⁸, M. Wallraff⁷, M. Walter⁴, R. Wasserman¹⁸, Ch. Weaver¹, C. Wendt¹, S. Westerhoff¹, N. Whitehorn¹, K. Wiebe³³, C. H. Wiebusch², D. R. Williams⁴⁷, R. Wischnewski⁴, H. Wissing¹⁹, M. Wolf²¹, T. R. Wood³², K. Woschnagg¹¹, C. Xu⁸, D. L. Xu⁴⁷, X. W. Xu³⁰, J. P. Yanez⁴, G. Yodh¹⁰, S. Yoshida³⁶, P. Zarzhitsky⁴⁷ & M. Zoll²¹

¹Department of Physics, University of Wisconsin, Madison, Wisconsin 53706, USA.

²Department of Physics and Astronomy, University of Gent, B-9000 Gent, Belgium.

³Department of Physics, University of Wisconsin, River Falls, Wisconsin 54022, USA.

⁴DESY, D-15735 Zeuthen, Germany. ⁵Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. ⁶Département de physique nucléaire et corpusculaire, Université de Genève, CH-1211 Genève, Switzerland. ⁷Ill. Physikalisches Institut, RWTH Aachen University, D-52056 Aachen, Germany. ⁸Bartol Research Institute and Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA. ⁹Department of Physics, South Dakota School of Mines and Technology, Rapid City, South Dakota 57701, USA. ¹⁰Department of Physics and Astronomy, University of California, Irvine, California 92697, USA. ¹¹Department of Physics, University of California, Berkeley, California 94720, USA. ¹²Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ¹³Department of Physics and

Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, Ohio 43210, USA. ¹⁴Department of Astronomy, Ohio State University, Columbus, Ohio 43210, USA. ¹⁵Université Libre de Bruxelles, Science Faculty CP230, B-1050 Brussels, Belgium.

¹⁶Fakultät für Physik & Astronomie, Ruhr-Universität Bochum, D-44780 Bochum, Germany.

¹⁷Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany.

¹⁸Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

¹⁹Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

²⁰Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA.

²¹Oskar Klein Centre and Department of Physics, Stockholm University, SE-10691 Stockholm, Sweden.

²²Vrije Universiteit Brussel, Dienst ELEM, B-1050 Brussels, Belgium.

²³Physikalisches Institut, Universität Bonn, Nussallee 12, D-53115 Bonn, Germany.

²⁴Department of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden.

²⁵Department of Physics, TU Dortmund University, D-44221 Dortmund, Germany.

²⁶Laboratory for High Energy Physics, École Polytechnique Fédérale, CH-1015 Lausanne, Switzerland.

²⁷Max-Planck-Institut für Kernphysik, D-69177 Heidelberg, Germany.

²⁸Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

²⁹School of Physics and Center for Relativistic Astrophysics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

³⁰Department of Physics, Southern University, Baton Rouge, Louisiana 70813, USA.

³¹Department of Astronomy, University of Wisconsin, Madison, Wisconsin 53706, USA.

³²Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2G7.

³³Institute of Physics, University of Mainz, Staudinger Weg 7, D-55099 Mainz, Germany.

³⁴School of Chemistry & Physics, University of Adelaide, Adelaide 5005, South Australia, Australia.

³⁵Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

³⁶Department of Physics, Chiba University, Chiba 263-8522, Japan.

³⁷CTSPS, Clark-Atlanta University, Atlanta, Georgia 30314, USA.

³⁸Institut für Physik, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany.

³⁹Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA.

⁴⁰Université de Mons, 7000 Mons, Belgium.

⁴¹Dipartimento di Fisica, Sezione INFN, I-70126, Bari, Italy.

⁴²Department of Physics and Astronomy, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, Alaska 99508, USA.

⁴³Technische Universität München, D-85748 Garching, Germany.

⁴⁴Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

⁴⁵Department of Physics, University of the West Indies, Cave Hill Campus, Bridgetown BB11000, Barbados.

⁴⁶NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

⁴⁷Department of Physics and Astronomy, University of Alabama, Tuscaloosa, Alabama 35487, USA.

Coherent quantum phase slip

O. V. Astafiev^{1,2}, L. B. Ioffe³, S. Kafanov^{1,2}, Yu. A. Pashkin^{1,2,4}, K. Yu. Arutyunov^{5,6}, D. Shahr⁷, O. Cohen⁷ & J. S. Tsai^{1,2}

A hundred years after the discovery of superconductivity, one fundamental prediction of the theory, coherent quantum phase slip (CQPS), has not been observed. CQPS is a phenomenon exactly dual¹ to the Josephson effect; whereas the latter is a coherent transfer of charges between superconducting leads^{2,3}, the former is a coherent transfer of vortices or fluxes across a superconducting wire. In contrast to previously reported observations^{4–8} of incoherent phase slip, CQPS has been only a subject of theoretical study^{9–12}. Its experimental demonstration is made difficult by quasiparticle dissipation due to gapless excitations in nanowires or in vortex cores. This difficulty might be overcome by using certain strongly disordered superconductors near the superconductor–insulator transition. Here we report direct observation of CQPS in a narrow segment of a superconducting loop made of strongly disordered indium oxide; the effect is made manifest through the superposition of quantum states with different numbers of flux quanta¹³. As with the Josephson effect, our observation should lead to new applications in superconducting electronics and quantum metrology^{1,10,11}.

Phase slips across superconducting wires¹⁴ lead to non-zero resistance^{2,15}; they may also lead to qubit dephasing, as result of charge noise¹⁶ in a chain of Josephson junctions¹⁷. Resistance measurements are dissipative, so its saturation at low temperatures cannot be interpreted as evidence for coherent QPS; however, a blockade of supercurrent might indicate CQPS, as demonstrated in ref. 8. A system decoupled from the environment has been suggested¹³: a superconducting loop in which phase slips change the number of quantized fluxes¹⁸, resulting in their quantum superposition, therefore, exhibiting CQPS⁹. We report a successful implementation of this idea using superconducting loops made of highly disordered InO_x.

We begin by considering a superconducting loop of effective area S with a high kinetic inductance L_k , containing a narrow segment (nanowire) with a finite rate of QPS. The external magnetic field B_{ext} perpendicular to the loop induces a flux $\Phi_{\text{ext}} = B_{\text{ext}}S$. The states of the loop are described by the phase winding number, N ; their energies are $E_N = (\Phi_{\text{ext}} - N\Phi_0)^2/2L_k$ (Fig. 1a). The energy difference between adjacent states $|N+1\rangle$ and $|N\rangle$ is $E_{N+1} - E_N = 2I_p\delta\Phi$, where $\delta\Phi = \Phi_{\text{ext}} - (N+1/2)\Phi_0$ and $I_p = \Phi_0/2L_k$ is the loop persistent current. At $\delta\Phi = 0$, the two states are degenerate. The QPS process, characterized by the amplitude E_S , couples the flux states, resulting in the Hamiltonian

$$H = -\frac{1}{2}E_S(|N+1\rangle\langle N| + |N\rangle\langle N+1|) + E_N|N\rangle\langle N| \quad (1)$$

which is dual to the Hamiltonian of a superconducting island connected to a reservoir through a Josephson junction. In the latter, L_k is replaced by capacitance, E_S by the Josephson energy, and N is the number of the Cooper pairs on the island¹³. The energy splitting between the ground and excited states $|g\rangle = \sin\frac{\theta}{2}|N\rangle + \cos\frac{\theta}{2}|N+1\rangle$ and $|e\rangle = \cos\frac{\theta}{2}|N\rangle - \sin\frac{\theta}{2}|N+1\rangle$ is $\Delta E = \sqrt{(2I_p\delta\Phi)^2 + E_S^2}$, where the mixing angle is $\theta = \arctan[E_S/(2I_p\delta\Phi)]$.

To detect CQPS, the loop is coupled to the coplanar line (resonator) by mutual inductance, M (ref. 19). In a rotating wave approximation, the effective Hamiltonian of the system resonantly driven by a classical microwave field with current amplitude $I_{\text{mw}} \cos(\Delta Et/\hbar)$ is $H_{\text{rw}} = \frac{\hbar\Omega}{2}(|e\rangle\langle g| + |g\rangle\langle e|)$, where $\hbar\Omega = MI_p I_{\text{mw}} \frac{E_S}{\Delta E}$ and \hbar is the reduced Planck's constant, $\hbar/2\pi$. Note that a transition between the two states can be induced only when $E_S \neq 0$, so observation of the spectroscopy peak constitutes direct evidence for CQPS.

Now we provide theoretical background for our choice of material. CQPS results from quantum fluctuations of the order parameter. Generally, the impact of fluctuations is characterized by the Ginzburg parameter, G_i , which is essentially the inverse number of Cooper pairs in a volume ξ^3 : $G_i = (\xi^3 v \Delta)^{-1}$, where v is the electron density of states, Δ is the superconducting gap and ξ is the coherence length. Even in disordered bulk superconductors characterized by $k_F l \approx 1$ (where k_F is the Fermi wavevector and l is the mean free path), G_i is small: $G_i \approx (k_F \xi)^{-1} \ll 1$. Strong fluctuations require materials with an even higher degree of disorder, in which electrons are localized. The superconductivity is suppressed if the localization length, ξ_{loc} , becomes shorter than ξ . Following this reasoning, we expect that the fluctuations are maximal ($G_i \approx 1$), when $\xi_{\text{loc}} \approx \xi$, that is, close to the superconductor–insulator transition. Although BCS theory² fails to describe this transition, it provides a qualitatively correct explanation for the behaviour of materials with $\xi_{\text{loc}} \approx \xi$. The fluctuations are

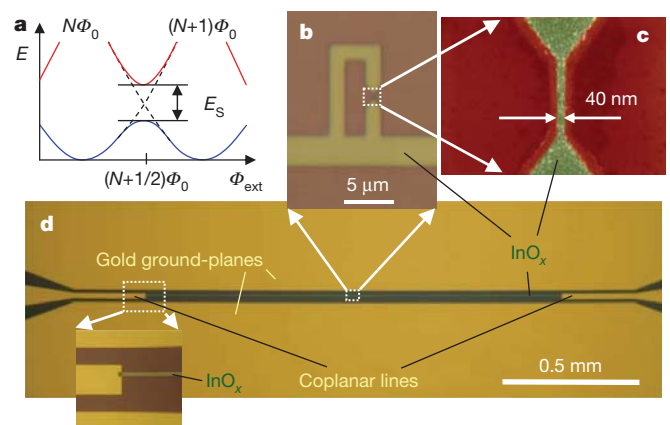


Figure 1 | The device. **a**, Energies of the loop versus external flux, Φ_{ext} . Blue and red lines: ground and excited energy levels, respectively. The degeneracy between states with N and $N+1$ flux quanta (Φ_0), seen at $\Phi_{\text{ext}} = (N+1/2)\Phi_0$, is lifted by the phase-slip energy, E_S . **b**, InO_x loop with a narrow wire segment on the right side is attached to the resonator (horizontal line) at the bottom. **c**, False-colour scanning-electron micrograph of the narrow InO_x segment. **d**, Step-impedance resonator, comprising a 3-μm-wide InO_x strip with wave impedance $Z_1 \approx 1,600 \Omega$ galvanically coupled to a gold coplanar line with impedance $Z_0 = 50 \Omega$. The boundaries of the resonator are defined by the strong impedance mismatch ($Z_1 \gg Z_0$).

¹NEC Green Innovation Research Laboratories, 34 Miyukigaoka, Tsukuba, Ibaraki, 305-8501, Japan. ²The Institute of Physical and Chemical Research (RIKEN), 34 Miyukigaoka, Tsukuba, Ibaraki, 305-8501, Japan. ³Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, New Jersey 08854, USA. ⁴Department of Physics, Lancaster University, Lancaster LA1 4YB, UK. ⁵University of Jyväskylä, Department of Physics, PB 35, 40014 Jyväskylä, Finland. ⁶Moscow State University, Institute of Nuclear Physics, Leninskiiye gory, GSP-1, Moscow 119899, Russia. ⁷Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

enhanced in narrow wires with a small number of effective conductive channels, $N_{\text{ch}} = R_K/R_\xi$, where R_ξ is the resistance of a wire of length ξ and $R_K = (h/e^2)$. The phase-slip amplitude decreases with N_{ch} as $E_s \propto \exp(-aN_{\text{ch}})$ (refs 20, 21), where a is a numerical parameter of order 1. Our data are consistent with $a \approx 0.3$ for $R_\xi = 1 \text{ k}\Omega$ and $\xi = 10 \text{ nm}$ (Supplementary Information). Thus, E_s is expected to be sizeable in strongly disordered, quasi-one-dimensional superconducting wires. But a high degree of disorder may also enhance Coulomb repulsion between electrons, and turn the superconductor into a dissipative normal metal²². The ideal materials for observation of CQPS should therefore form localized, pre-formed pairs even before they become superconducting^{23–25}. Such behaviour has been observed in amorphous InO_x and TiN thin films^{26,27}.

Our loops (Fig. 1b) are fabricated from a 35-nm-thick superconducting InO_x film with $\xi = 10\text{--}30 \text{ nm}$ (ref. 28), using electron-beam lithography. The uncertainty in the magnitude of ξ is due to the poor applicability of BCS theory (see Supplementary Information). Each loop contains a narrow wire segment, about 40 nm wide and 400 nm long (Fig. 1c). This wire is reasonably homogeneous, varying in width by less than 10 nm. We have characterized identical wires and films, fabricated by the same process, using d.c. transport measurements. The film exhibits a superconducting transition at $T_c = 2.7 \text{ K}$. The sheet resistance R_{sq} of the wide films, measured slightly above T_c , is 1.7 k Ω , which gives a rough (BCS) estimate of the sheet inductance as $L_{\text{sq}} \approx \hbar R_{\text{sq}}/\pi A \approx 0.7 \text{ nH}$ (ref. 13) (for $A \approx 0.5 \text{ meV}$; ref. 26) for the wide sections of the loops.

To measure the loops, we incorporate them in a step-impedance coplanar resonator: a strip line of InO_x of length $L = 1.5 \text{ mm}$ and width $W = 3 \mu\text{m}$, galvanically connected to two gold coplanar lines (with impedance $Z_0 = 50 \Omega$) at the ends (Fig. 1d). We estimate the effective wave impedance of the line to be $Z_1 = \sqrt{l_1/c_1} \approx 1.6 \text{ k}\Omega$, with specific inductance $l_1 = L_{\text{sq}}/W \approx 2.3 \times 10^{-4} \text{ H m}^{-1}$ and specific capacitance $c_1 \approx 0.85 \times 10^{-10} \text{ F m}^{-1}$. Because $Z_1 \gg Z_0$, the strong mismatch results in standing wave formation with maximal current at the boundaries: $I(x) = I_m \cos(\pi m x/L)$, where x is the coordinate along the resonator ($0 \leq x \leq L$). The resonance frequency of the m th mode f_m ($m = 1, 2, 3, \dots$) is given by $f_m \approx mv/2L = m \times 2.4 \text{ GHz}$, where $v = (l_1 c_1)^{-1/2} \approx 7.2 \times 10^6 \text{ m s}^{-1}$ is the group velocity in the resonator. The energy decay rate in such a resonator is $\kappa \approx (4Z_0/Z_1)(v/L)$, which limits the power peak width to $\kappa/2\pi \approx 0.1 \text{ GHz}$.

Our main results are as follows. Figure 2a shows the power transmission coefficient $|t|^2$ through the resonator at a temperature of 40 mK. The peaks correspond to $m = 3, 4, 5$, with resonance frequencies $f_3 = 6.65 \text{ GHz}$, $f_4 = 9.08 \text{ GHz}$ and $f_5 = 11.00 \text{ GHz}$ (close to our estimates above). The actual peak widths (full-width at half-maximum) are approximately 250 MHz ($= \kappa/2\pi$), which is twice as large as that expected from the loading loss. This is probably due to an extra loss in the gold ground-plane films. Our loops located in the middle of the resonator ($x = L/2$) are coupled only to the odd modes m , for which the current defined by $I_{\text{mw}} = I_m \cos(\pi m/2)$ is non-zero.

To detect the superposition of flux states we measure transmission t through the resonator at f_m versus B_{ext} . The transmission does not depend on B_{ext} at the third and fifth peaks. However, at the fourth mode peak, t exhibits well pronounced periodic structure: sharp negative dips in the amplitude $|t|$, and in the phase rotation $\arg(t)$, as shown in Fig. 2b. The period, $\Delta B = 0.061 \text{ mT}$, corresponds with high accuracy to one flux quantum, Φ_0 , through the area $S = 32 \mu\text{m}^2$ of the loop shown in Fig. 1b, with a 40-nm-wide wire. Note that there is a small ($\sim 5\%$) uncertainty in the loop area, owing to the finite width of the lines. Each resonator contains five loops, with nanowire widths 40, 60, 80, 100 and 120 nm. The loops differ slightly in area, varying by 10–12% from one to another. No CQPS-related signals are found from the samples with nanowire segments wider than 40 nm, indicating that in these loops $E_s/\hbar < 1 \text{ GHz}$, in good agreement with our prediction that E_s will be suppressed by more than a factor of 10 with an increase of 20 nm in the nanowire width. (See Supplementary Information.)

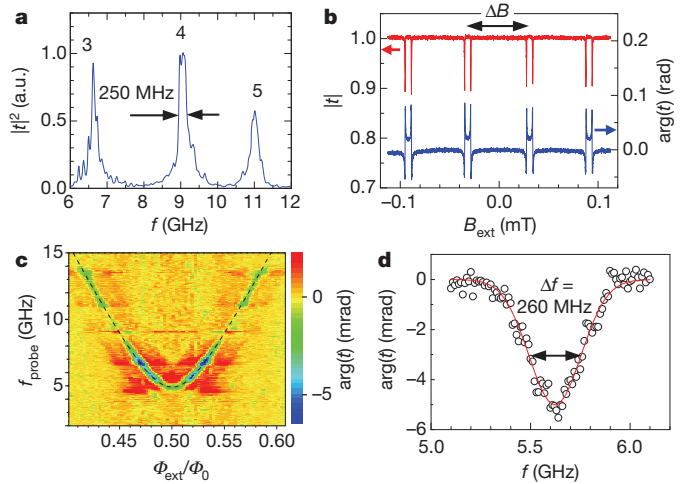


Figure 2 | Experimental data. **a**, Power transmission through the resonator, measured within the bandwidth of our experimental setup. Peaks in transmission power coefficient, $|t|^2$, correspond to resonator modes, with mode number m indicated for each peak. a.u., arbitrary units. **b**, Transmission through the resonator as function of external magnetic field B_{ext} at $m = 4$ ($f_4 = 9.08 \text{ GHz}$). The periodic structure in amplitude ($|t|$) and phase ($\arg(t)$) corresponds to the points where the lowest-level energy gap $\Delta E/\hbar$ matches f_4 . The period $\Delta B = 0.061 \text{ mT}$ ($= \Phi_0/S$) indicates that the response comes from the loop (shown in Fig. 1b), with the effective loop area $S = 32 \mu\text{m}^2$. **c**, The two-level spectroscopy line obtained in two-tone measurements. The phase of transmission, $\arg(t)$, through the resonator at f_4 is monitored, while another tone with frequency f_{probe} from an additional microwave generator, and B_{ext} , are independently swept. The plot is filtered to eliminate the contribution of other resonances ($2 \leq m \leq 6$), visible as horizontal red features. The dashed line is the fit to the energy splitting, with $\Delta E/\hbar = 4.9 \text{ GHz}$, $I_p = 24 \text{ nA}$. **d**, The resonant dip is measured at $\Phi/\Phi_0 = 0.52$. The red curve is the Gaussian fit.

Although the coupling of our loop to the resonator ($g/\hbar \approx 10 \text{ MHz}$) is weak ($g < \hbar\kappa$), we were able to perform spectroscopy measurements by monitoring resonator transmission^{29,30} while scanning B_{ext} and the frequency, f_{probe} , of an additional probe microwave tone. The transmission phase plot shows the resonance excitation of the two-level system (Fig. 2c). The obtained structure is distorted by periodic resonances at f_m (seen as red horizontal features), and the resulting picture is plotted after filtering out these resonances. The green–blue line corresponds to the expected energy splitting, which is well fit by $\Delta E = \sqrt{(2I_p \delta\Phi)^2 + E_s^2}$ (dashed line), with the fitting parameters $E_s/\hbar < 4.9 \text{ GHz}$ and $I_p = 24 \text{ nA}$, corresponding to $L_k = 42 \text{ nH}$.

The exact splitting at the degeneracy point $\delta\Phi = 0$ happens to nearly coincide with $f_2 \approx 4.8 \text{ GHz}$, distorting the line shape. Figure 2d shows the resonant peak at $\Phi_{\text{ext}} = 0.52 \Phi_0$, slightly away from the degeneracy point, where the resonant frequency $\Delta E/\hbar \approx 5.6 \text{ GHz}$ is between f_1 and f_2 . The spectroscopy peak is well fit by a Gaussian, $A \exp\left[-\frac{1}{2}\left(\frac{f - \Delta E/\hbar}{\Delta f}\right)^2\right]$, with $\Delta f = 260 \text{ MHz}$.

This demonstrates coherent coupling between the flux states in the loop. An interesting question, which requires further study, is the mechanism of decoherence in this type of system. At this stage, we can only conjecture that the shape suggests peak broadening due to low-frequency Gaussian noise, rather than relaxation.

We now discuss the relation between the measured and expected properties of the system. The total loop resistance R (at $T > T_c$) is the sum of the resistances of the wide section (estimated from R_{sq}) and of the nanowire, $R_w \approx 30 \text{ k}\Omega$, deduced from the d.c. measurements of identical samples. We estimate $R \approx 55 \text{ k}\Omega$. As one might expect, the effective sheet resistance of the narrow wires is higher than R_{sq} —in our case, by a factor of 1.8. The BCS estimate of the inductance of the whole loop gives $L_k \approx 23 \text{ nH}$. Close to the superconductor–insulator transition

(SIT), R rises continuously with decreasing temperature before the onset of superconductivity, whereas L_k diverges at the transition. Arguing by continuity, BCS theory should underestimate L_k in the strongly disordered superconductors (see Supplementary Information). We underestimate L_k in our films by a factor of 1.8, which is very reasonable for the superconductors close to the SIT.

Another measure of the wire quality is the value of its critical current. For a BCS superconductor, the maximal supercurrent¹³ is $I_c \approx 0.75 \Delta / e R_\xi$ (Supplementary Information). Taking into account that in the vicinity of the SIT, I_c is overestimated by at least the same factor of 1.8 (as with L_k), we find $I_c \leq 280$ –90 nA, for $\xi = 10$ –30 nm. The measured value, $I_c = 100$ nA, is in full agreement with these expectations.

The quantitative estimate of the QPS amplitude is less accurate than for kinetic inductance or critical current, owing to its exponential dependence on sample and material parameters. Moreover, in the BCS approximation, QPS amplitude explicitly depends on ξ , which is not well defined close to the SIT. Rough estimates based on ref. 15 give values of ~ 5 –20 GHz for $\xi = 10$ nm; our calculations, based on the measured value of the sheet kinetic inductance, yield $E_S \approx 10$ GHz, in agreement with the data (see Supplementary Information).

We have obtained additional evidence for CQPS, as follows. In addition to the data presented here for a single sample, we observed very similar results in two other samples fabricated from the same InO_x film. One of those samples was measured in the step-impedance resonator described above, and the other in an open-line configuration¹⁹. In all three samples, we found resonances with close values of persistent current, and level splitting E_S/h equal to 4.9, 5.8 and 9.5 GHz, respectively. The variation in E_S is surprisingly low (given its exponential dependence on parameters), and is in agreement with expectations (see

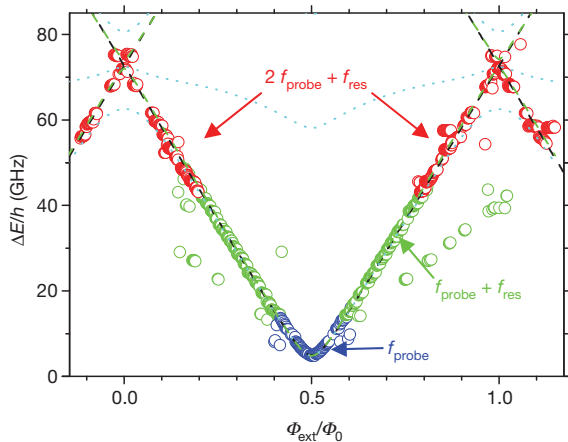


Figure 3 | Spectroscopy of the system across a wide range of flux and frequency. The spectroscopy response is obtained by measuring the transmission amplitude and phase through the resonator at $f_4 = 9.08$ GHz, while the frequency of the additional microwave tone f_{probe} is swept. Depending on the range, the two-level system excitation frequency $\Delta E/h$ is derived from the direct (single-photon) excitation $\Delta E/h = f_{\text{probe}}$ (blue circles), the two-photon process $\Delta E/h = f_{\text{probe}} + f_4$ (green circles), or the three-photon process $\Delta E/h = 2f_{\text{probe}} + f_4$ (red circles). With these methods, using $f_{\text{probe}} \leq 35$ GHz, we could trace $\Delta E/h$ up to about 77 GHz. The dashed black line is the calculated energy splitting, $\Delta E/h = \sqrt{(2I_p \delta \Phi)^2 + E_S^2} / h$, with $E_S = 4.9$ GHz and $I_p = 24$ nA. Perfect agreement of the experimental data with the calculated energy bands supports our interpretation; the quantum states of the system correspond to the superposition of the two adjacent flux states induced by CQPS. The rare scattered points result from noise, resonator resonances and higher-order excitation processes. For comparison, the additional dotted cyan lines show the expected behaviour of a radio-frequency-SQUID qubit with $E_J/h = 50$ GHz ($I_c = 100$ nA), capacitance $C = 1.1$ fF and inductance $L = 38$ nH, which qualitatively disagrees with our data at the degeneracy points $\Phi_{\text{ext}} = 0$ and $\Phi_{\text{ext}} = \Phi_0$.

Supplementary Information). This reproducibility is a strong argument against an alternative interpretation of the results based on unintentional formation of a rogue Josephson junction somewhere in the wire. Second, as in a charge qubit (which is dual to our system with CQPS), the energy difference between the two lowest states in the loop asymptotically approaches linear dependence, $\Delta E = 2I_p \delta \Phi$, at $\Delta E \gg E_S$. In our experiments, we are able to trace the two-level system resonance up to $|\delta \Phi| \geq 0.5 \Phi_0$ and $\Delta E \approx 77$ GHz (Fig. 3). This linear dependence, without any observable splitting at $\Phi = 0$ and Φ_0 (corresponding to the second-order 4π -process), indicates a linear inductance in our system, rather than a nonlinear one originating in a rogue Josephson junction. To support this statement, we present simulations of the spectroscopy lines (dotted cyan lines in Fig. 3) for a radio-frequency-SQUID qubit with a single Josephson junction (with $I_c = 100$ nA), and with capacitance and linear inductance adjusted to provide the best fit of our spectroscopy data. The simulations show a significant rounding at $\Phi = 0$ and Φ_0 , owing to the nonlinear Josephson inductance, in qualitative disagreement with our results. These arguments allow us to exclude a rogue Josephson junction formed in our device; however, we cannot distinguish a number of interfering phase slip locations in the wire, if, for example, the wire is not uniform.

Received 24 October 2011; accepted 3 February 2012.

1. Mooij, J. E. & Nazarov, Yu. V. Superconducting nanowires as quantum phase-slip junctions. *Nature Phys.* **2**, 169–172 (2006).
2. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, 1996).
3. Averin, D. V., Zorin, A. B. & Likharev, K. K. Bloch oscillations in small Josephson junction. *Zh. Eksp. Teor. Fiz.* **88**, 407–412 (1984).
4. Giordano, N. Evidence for macroscopic quantum tunnelling in one-dimensional superconductors. *Phys. Rev. Lett.* **61**, 2137–2140 (1988).
5. Bezryadin, A., Lau, C. N. & Tinkham, M. Quantum suppression of superconductivity in ultrathin nanowires. *Nature* **404**, 971–974 (2000).
6. Zgirski, M., Riikonen, K.-P., Touboltsev, V. & Arutyunov, K. Yu. Quantum fluctuations in ultranarrow superconducting aluminum nanowires. *Phys. Rev. B* **77**, 054508 (2008).
7. Lehtinen, J. S., Sajavaara, T., Arutyunov, K., Yu. & Vasiliev, A. Evidence of quantum phase slip effect in titanium nanowires. Preprint at (<http://arxiv.org/abs/1106.3852>) (2011).
8. Hongisto, T. T. & Zorin, A. B. Single charge transistor based on superconducting nanowire in high impedance environment. *Phys. Rev. Lett.* **108**, 097001 (2012).
9. Matveev, K. A., Larkin, A. I. & Glazman, L. I. Persistent current in superconducting nanorings. *Phys. Rev. Lett.* **89**, 096802 (2002).
10. Hrisu, A. M. & Nazarov, Yu. V. Model of a proposed superconducting phase slip oscillator: A method for obtaining few-photon nonlinearities. *Phys. Rev. Lett.* **106**, 077004 (2011).
11. Hrisu, A. M. & Nazarov, Yu. V. Coulomb blockade due to quantum phase-slips illustrated with devices. *Phys. Rev. B* **83**, 174511 (2011).
12. Vanevic, M. & Nazarov, Yu. V. Quantum phase slips in superconducting wires with weak links. Preprint at (<http://arxiv.org/abs/1108.3553>) (2011).
13. Mooij, J. E. & Harmans, C. J. P. M. Phase-slip flux qubits. *N. J. Phys.* **7**, 219 (2005).
14. Little, W. A. Decay of persistent current in small superconductors. *Phys. Rev.* **156**, 396–403 (1967).
15. Arutyunov, K. Yu., Golubev, D. S. & Zaikin, A. D. Superconductivity in one dimension. *Phys. Rep.* **464**, 1–70 (2008).
16. Manucharyan, V. E. et al., *Phys. Rev. B* **85**, 024521 (2012).
17. Pop, I. M. et al. Experimental demonstration of Aharonov-Casher interference in a Josephson junction circuit. Preprint at (<http://arxiv.org/abs/1104.3999>) (2011).
18. Arutyunov, K. Yu., Hongisto, T. T., Lehtinen, J. S., Leino, L. I. & Vasiliev, A. L. Quantum phase-slip phenomenon in ultra-narrow superconducting nanorings. *Sci. Rep.* **2**, 293 (2012).
19. Astafiev, O. et al. Resonance fluorescence of a single artificial atom. *Science* **327**, 840–843 (2010).
20. Zaikin, A. D., Golubev, D. S., van Otterlo, A. & Zimanyi, G. T. Quantum phase slips and transport in ultrathin superconducting wires. *Phys. Rev. Lett.* **78**, 1552–1555 (1997).
21. Golubev, D. S. & Zaikin, A. D. Quantum tunneling of the order parameter in superconducting nanowires. *Phys. Rev. B* **64**, 014504 (2001).
22. Finkel'stein, A. M. Suppression of superconductivity in homogeneously disordered systems. *Physica B* **197**, 636–648 (1994).
23. Feigel'man, M. V., Ioffe, L. B., Kravtsov, V. E. & Cuevas, E. Fractal superconductivity near localization threshold. *Ann. Phys.* **325**, 1390–1478 (2010).
24. Feigel'man, M. V., Ioffe, L. B., Kravtsov, V. E. & Yuzbashyan, E. A. Eigenfunction fractality and pseudogap state near the superconductor-insulator transition. *Phys. Rev. Lett.* **98**, 027001 (2007).
25. Feigel'man, M. V., Ioffe, L. B. & Mezard, M. Superconductor-insulator transition and energy localization. *Phys. Rev. B* **82**, 184534 (2010).

26. Sacépé, B. *et al.* Localization of preformed Cooper pairs in disordered superconductors. *Nature Phys.* **7**, 239–244 (2011).
27. Sacépé, B. *et al.* Pseudogap in a thin film of a conventional superconductor. *Nature Commun.* **1**, 140 (2010).
28. Johansson, A., Sambandamurthy, G., Shahar, D., Jacobson, N. & Tenne, R. Nanowire acting as superconducting quantum device. *Phys. Rev. Lett.* **95**, 116805 (2005).
29. Wallraff, A. *et al.* Approaching unit visibility for control of a superconducting qubit with dispersive readout. *Phys. Rev. Lett.* **95**, 060501 (2005).
30. Abdumalikov, A. A. Jr, Astafiev, O. V., Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. Vacuum Rabi splitting due to strong coupling of a flux qubit and a coplanar-waveguide resonator. *Phys. Rev. B* **78**, 180502 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Feigel'man, J. Mooij and Y. Nazarov for discussions. This work was supported by Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST), MEXT KAKENHI “Quantum Cybernetics”, Ministry of Science and Education of Russian Federation grant

2010-1.5-508-005-037. L.B.I. was supported by ARO W911NF-09-1-0395, DARPA HR0011-09-1- 0009 and NIRT ECS-0608842. D.S. and O.C. were supported by Minerva Fund.

Author Contributions O.V.A. planned the experiment, designed and fabricated the samples, performed measurements and data analysis. L.B.I. came up with the idea of using materials close to the SIT and provided theoretical support. S.K. fabricated the sample and contributed to understanding the data. Yu.A.P. participated in discussions of the experiment. K.Yu.A. triggered the research direction and suggested the realization of the phase-slip qubit. D.S. and O.C. fabricated the InO_x films. J.S.T. discussed the data and provided support for the work within the FIRST and KAKENHI projects. O.V.A. wrote the manuscript with feedback from all authors, including significant contributions from L.B.I. and K.Yu.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.V.A. (astf@zb.jp.nec.com).

Air density 2.7 billion years ago limited to less than twice modern levels by fossil raindrop imprints

Sanjoy M. Som^{1,2†}, David C. Catling¹, Jelte P. Harnmeijer^{1,3}, Peter M. Polivka^{1,4} & Roger Buick¹

According to the ‘Faint Young Sun’ paradox, during the late Archaean eon a Sun approximately 20% dimmer warmed the early Earth such that it had liquid water and a clement climate¹. Explanations for this phenomenon have invoked a denser atmosphere that provided warmth by nitrogen pressure broadening¹ or enhanced greenhouse gas concentrations². Such solutions are allowed by geochemical studies and numerical investigations that place approximate concentration limits on Archaean atmospheric gases, including methane, carbon dioxide and oxygen^{2–7}. But no field data constraining ground-level air density and barometric pressure have been reported, leaving the plausibility of these various hypotheses in doubt. Here we show that raindrop imprints in tuffs of the Ventersdorp Supergroup, South Africa, constrain surface air density 2.7 billion years ago to less than twice modern levels. We interpret the raindrop fossils using experiments in which water droplets of known size fall at terminal velocity into fresh and weathered volcanic ash, thus defining a relationship between imprint size and raindrop impact momentum. Fragmentation following raindrop flattening limits raindrop size to a maximum value independent of air density, whereas raindrop terminal velocity varies as the inverse of the square root of air density. If the Archaean raindrops reached the modern maximum measured size, air density must have been less than 2.3 kg m^{−3}, compared to today’s 1.2 kg m^{−3}, but because such drops rarely occur, air density was more probably below 1.3 kg m^{−3}. The upper estimate for air density renders the pressure broadening explanation¹ possible, but it is improbable under the likely lower estimates. Our results also disallow the extreme CO₂ levels required for hot Archaean climates⁸.

Numerical investigations of Archaean atmospheric composition^{2–6} typically assume a modern, total atmospheric pressure of about one atmosphere (1 atm), but there are good reasons why barometric pressure may have been different. First, the partial pressure of oxygen p_{O_2} was negligible before the Great Oxidation Event at around 2.35 billion years ago⁹. There are several independent lines of evidence for this¹⁰, the strongest being widespread and large mass-independent fractionations of sulphur isotopes in Archaean sediments that arise only in an atmosphere with less than about one part oxygen per million by volume (p.p.m.v.)¹¹. Second, before the advent of an aerobic nitrogen cycle coincident with rising oxygen levels¹², the flux of nitrogen back to the atmosphere via the now-dominant nitrification–denitrification pathway would have been different from now. So a lack of oxygen before the Great Oxidation Event should have affected the partial pressure of nitrogen p_{N_2} , the major gas contributing to total atmospheric pressure. Moreover, it has been calculated that a p_{N_2} of 2.37 atm at 2.5 billion years ago could solve the ‘Faint Young Sun’ paradox by pressure-broadening infrared absorption of greenhouse gases¹. Other studies postulate a hot (~70 °C) Archaean ocean based on oxygen isotopes in cherts¹³, requiring a partial pressure of carbon dioxide p_{CO_2} of about 2–6 bar (ref. 8), which would contradict the p_{CO_2} levels of only 10–50 present atmospheric levels (PAL) constrained

from 2.69-billion-year-old palaeosols⁷. Such ambiguities concerning the composition of the ancient atmosphere could be resolved, or improved upon, by knowledge of total atmospheric pressure. Here, we use raindrop imprints to constrain total ground-level atmospheric density (and thus total surface pressure) 2.7 billion years before present. The idea of using raindrop imprints as a proxy for air density was suggested by Lyell¹⁴ in 1851 but has hitherto been unexplored.

On the ancient Earth, maximum raindrop diameters should have been essentially identical to today’s, because the maximum size beyond which raindrops disintegrate at terminal velocity is independent of air density. Falling raindrops flatten^{15,16} and fragment when the total aerodynamic forces exceed the combination of surface tension and hydrostatic forces¹⁷. Fragmentation begins when the raindrop bottom becomes flat at a force balance given by¹⁷:

$$V_{\text{term}}^2 d = \frac{8\gamma}{\rho_{\text{air}} n C_d} \quad (1)$$

where V_{term} is terminal velocity, d is the diameter of a sphere equivalent to the drop volume, γ is surface tension, ρ_{air} is air density, C_d is the drag coefficient, and n is a factor relating the radius of the upper curvature of the drop to its spherical equivalent radius. Theory relates terminal velocity to raindrop size^{18,19} and predicts 9.3 m s^{−1} for a raindrop of 6.8 mm in diameter, the largest measured raindrop at ground level²⁰. Typical values under standard surface atmospheric conditions ($\gamma = 7 \times 10^{-2}$ N m^{−1} and $\rho_{\text{air}} = 1.2$ kg m^{−3}) yield a constant value of 0.80 for $n C_d$ in equation (1), comparing favourably with $n C_d = 0.85$ from independent studies¹⁷ and consistent with observations that the product $V_{\text{term}}^2 d$ is constant^{21,22}.

A further relationship derived from empirical correlations exists between air density and maximal terminal velocity^{18,23,24}:

$$V_{\text{term,max}} = 2 \left(\frac{\rho_{\text{water}} g \gamma}{\rho_{\text{air}}^2} \right)^{0.25} \approx \frac{\text{a known constant}}{\rho_{\text{air}}^{0.5}} \quad (2)$$

where g is gravitational acceleration and ρ_{water} is the density of water. Equation (2) also gives a maximum terminal velocity of 9.3 m s^{−1}, which corresponds to the largest raindrops of 6.8 mm in diameter. Substitution of $V_{\text{term}} \propto \rho_{\text{air}}^{-1/2}$ from equation (2) into $d \propto (V_{\text{term}}^2 \rho_{\text{air}})^{-1}$ from equation (1) cancels out ρ_{air} , showing that maximum raindrop size is independent of air density. Drop equivalent diameter d is thus simply a function of surface tension γ . The slight increase of surface tension with temperature causes only a trivial terminal velocity change of 0.05 m s^{−1} over 15–30 °C, meaning that somewhat different Archaean temperatures would not affect our conclusions. Consequently, an upper bound on air density can be derived from the largest raindrop imprints, formed by the transfer of momentum from the largest impacting raindrops to the substrate.

The Archaean imprints studied here (Fig. 1 and Supplementary Information) are from the Omdraaivlei farm near Prieska, South Africa, in the Kameeldoorns Formation of the Platberg Group, the middle unit of the 2.7-billion-year-old Ventersdorp Supergroup²⁵

¹Department of Earth and Space Sciences and Astrobiology Program, University of Washington, Seattle, Washington 98195-1310, USA. ²Blue Marble Space Institute of Science, Seattle, Washington 98145, USA. ³Sustainable Community Energy Network, Edinburgh Centre for Low Carbon Innovation, Edinburgh EH8 9AA, UK. ⁴Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, USA. [†]Present address: Exobiology Branch, NASA Ames Research Center, Moffett Field, California 94035, USA.

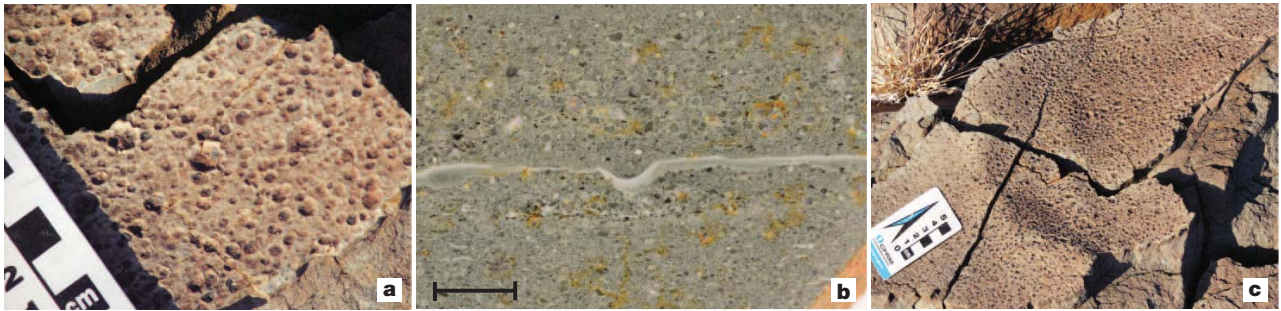


Figure 1 | The 2.7-billion-year-old Ventersdorp Supergroup raindrop imprints lithified in tuff at Omdraaivlei, South Africa. a, Detail of slightly elliptical outlines of raindrop imprints. **b,** Cross-section photograph of imprints penetrating 1–2 mm into coarse accretionary lapilli tuff, and draped

(Supplementary Fig. 2). They penetrate into very poorly sorted fine tuff of intermediate igneous composition. A layer of pale, very fine volcanic ash 0.5–0.8 mm thick drapes the imprints (Fig. 1b and Supplementary Information), which reduces the diameter of the exposed imprints relative to their original diameter by about twice the drape thickness. The rimmed craters are well preserved, slightly elliptical in shape and occasionally overlap, suggesting that the rain event that formed them was of short duration and of light to moderate intensity, because high-intensity rainfall leads to distorted imprints and long-duration showers cause substantial overlap²⁶. They were probably formed in an inland semi-arid setting near sea level (Supplementary Information).

The dominant contributor to imprint size is the change in raindrop momentum during impact²⁷. We define a dimensionless momentum J as:

$$J = \frac{V_{\text{term}} m_d}{\eta A_d} \quad (3)$$

where m_d is the mass of the raindrop, η is the dynamic viscosity (independent of ρ_{air}), and A_d is the cross-sectional area of the falling drop. Obtaining atmospheric density from lithified raindrop imprints requires: (1) measuring raindrop imprint areas; (2) determining experimentally how J varies with imprint area by varying d and thus m_d in equation (3) (Fig. 2); and (3) relating atmospheric density to J (Fig. 3). Archaean imprints were measured in the field, and subsequently re-measured by high-resolution three-dimensional laser scanning of latex peels taken *in situ*. The relationship between drop impact momentum and corresponding imprint area was obtained from experiments in which we released water drops of known mass from an indoor height sufficient to guarantee that they reached terminal velocity onto ash substrates analogous to the Archaean tuff. One of the experimental ashes was fresh from the 2010 Eyjafjallajökull

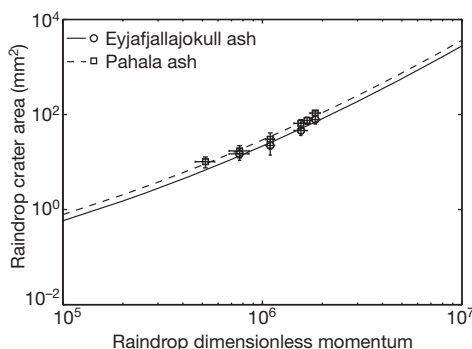


Figure 2 | Experimental relationship between raindrop area and dimensionless momentum. The horizontal error bars are the uncertainty of raindrop mass propagated to dimensionless momentum; the vertical error bars express the corresponding standard deviation of crater dimensions.

with a thin veneer (about 0.5 mm) of light-toned, fine volcanic ash. Scale bar, 10 mm. **c,** Mildly increased imprint density on the windward (north-facing) faces of underlying symmetrical wave-ripples. (Photo credits: W. Altermann for **a** and **c**, and T. Tobin for **b**).

eruption in Iceland, and the other was weathered late Pleistocene Pahala ash from Hawaii²⁸. Both were from mafic to intermediate hydro-volcanic eruptions with similar grain-size distributions to the ash hosting the Archaean raindrop imprints. The relationship between air density and dimensionless momentum was obtained by extrapolating from previous work relating air density to terminal velocity¹⁸ (Supplementary Information).

The actual raindrop diameter that formed the largest imprints found at Omdraaivlei is unknown. Smaller imprint area reflects lower raindrop velocity and thus higher air density. Figure 4 illustrates the expected atmospheric density when the raindrop size that caused the largest Omdraaivlei imprint is varied. To calculate an atmospheric density upper bound, we use the lower bound on the largest raindrop imprint area, because smaller imprint areas reflect lower raindrop terminal velocities and hence higher air density. The largest imprint area is bounded by $A_{\text{latex}} - 1\sigma$ and $A_{\text{in situ}} + 1\sigma$, where $A_{\text{in situ}}$ is the mean maximum imprint area measured in the field, and A_{latex} is the mean maximum imprint area measured from the latex peels (Supplementary Information). Finding the corresponding air density for these end-member dimensions for a fixed raindrop size defines the error in air density. The air density upper bound is calculated as $A_{\text{latex}} - 1\sigma$. Using a drop diameter of 6.8 mm—the size of the largest raindrop ever measured at ground level²⁰ and also the theoretical maximum size—we obtain an absolute upper limit of less than 2.3 kg m^{-3} . However, because rainfall events producing these maximal drops are extremely rare, very intense, and highly erosive²⁶, it is more likely that the maximum raindrop size responsible for creating the Archaean imprints had an equivalent diameter of 3.8–5.3 mm, depending upon the choice of parameterization of the raindrop size distribution and assuming that the probability distribution functions for rainfall rates in inland semi-arid settings were similar on the Archaean Earth and the modern Earth. These dimensions correspond to a more likely upper limit for atmospheric density of $0.6\text{--}1.3 \text{ kg m}^{-3}$.

Estimates of atmospheric pressure from the ideal gas law $P = \rho_{\text{air}} RT$ require assumptions about air temperature T and, through the specific gas constant R , atmospheric composition. Regarding temperature 2.7 ± 0.1 billion years ago, no evidence of glaciation is present in the rock record. This may reflect lack of preservation, but if it is real, an absence of glaciation requires average temperatures to have been 20°C or higher, according to data from non-glacial times in the Phanerozoic eon²⁹. This is also consistent with Archaean temperatures of less than 40°C based on oxygen isotope systematics³⁰. Taking a nominal temperature of about 20°C , we calculated an upper limit on atmospheric pressure by choosing a composition that maximizes R . A $100\% \text{ N}_2$ atmosphere ($R = 297 \text{ J kg}^{-1} \text{ K}^{-1}$, versus $R = 253 \text{ J kg}^{-1} \text{ K}^{-1}$ for a $70\% \text{ N}_2 + 30\% \text{ CO}_2$ atmosphere) constrains atmospheric pressure 2.7 billion years ago to below 0.52–1.1 atm if we take ρ_{air} as less than $0.6\text{--}1.3 \text{ kg m}^{-3}$ or an absolute upper limit of less than 2.1 atm if we take $\rho_{\text{air}} = 2.3 \text{ kg m}^{-3}$.

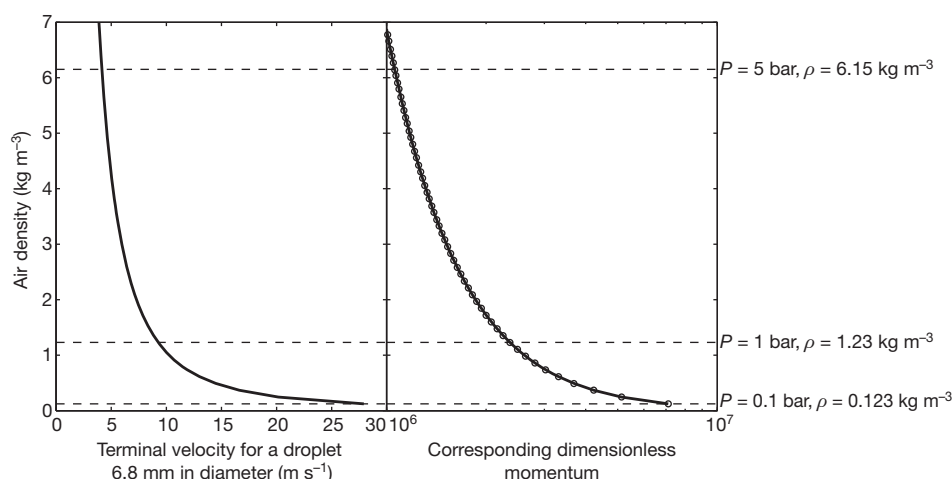


Figure 3 | Theoretical predictions of the variation of air density with terminal velocity and dimensionless momentum at the surface. **a**, The relationship between air density and terminal velocity, following the method of

ref. 18. **b**, The relationship between air density and dimensionless momentum (see Supplementary Information).

Our result extends quantifiable atmospheric pressure determinations beyond the modern era to the early Earth. It places constraints on some Archaean climate models^{1,13}, but does not invalidate other proposed late-Archaean atmospheres^{2–6}. For models invoking high p_{N_2} as a means of pressure-broadening greenhouse gas absorption¹, only the lowest estimate of 1.58 atm is close to our findings, suggesting a nitrogen content in the late-Archaean atmosphere of no more than twice present levels. Our result rules out very high Archaean ocean temperatures of 70 °C–85 °C (refs 13 and 31), because these would necessitate about 2–6 bar of carbon dioxide⁸ plus 0.3–0.6 bar of water vapour, increasing barometric pressure far beyond the upper limit found here. Thus, neither strong pressure-broadening nor extreme p_{CO_2} are satisfactory mechanisms for warming the early Earth illuminated by a ‘Faint Young Sun’.

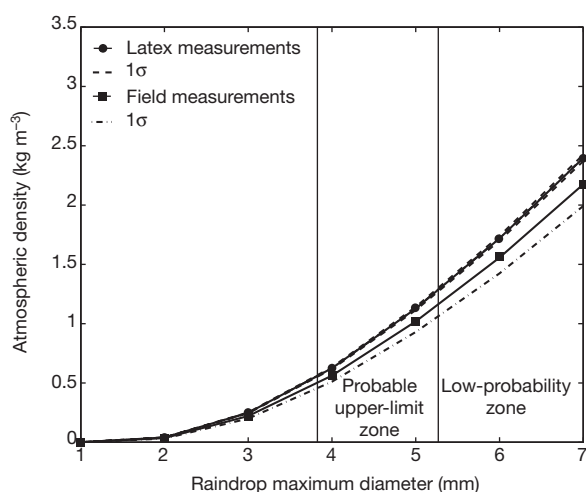


Figure 4 | Atmospheric density given the maximum raindrop diameter that created the largest imprints at Omdraaivlei, South Africa. Squares represent the air density calculated from the maximum imprint areas measured *in situ* (ash drape removed). Circles represent the air density calculated from the maximum imprint areas from latex measurement of imprint areas (ash drape removed). Dashed lines represent 1 σ error. With the assumption that the rainfall rate probability distribution function responsible for the Archaean imprints is analogous to a modern, semi-arid, rainfall rate probability distribution function, there is a 78–99% probability that the maximum raindrop diameter was less than 3.8–5.3 mm (Supplementary Fig. 5, Supplementary Table 1, and Supplementary Information).

METHODS SUMMARY

We measured the Ventersdorp raindrop imprints directly in the field and by casting them using low-viscosity latex for later laboratory study. The resulting latex peels captured the dimensions of 955 individual raindrop imprints. The topography of the peels was measured using high-resolution three-dimensional laser scanning. The corresponding point-clouds (available at http://gis.ess.washington.edu/papers/Sanjoy_Som_raindrops/) were interpolated using an inverse data-weighting scheme to create a digital elevation model. The digital elevation models were imported into a Geographical Information System and the dimensions of the imprints extracted. The dimensions were optimally binned, with the largest bin corresponding to the measurement of the largest imprints, and the dimension of each bin reflecting error in measurement.

To find the relationship between raindrop imprint dimension and dimensionless momentum, we released droplets of different (known) mass from a height of 27 m indoors onto analogous ash taken from Iceland and Hawaii. This height is more than double that required for all drops to reach terminal velocity. We could calculate the dimensionless momentum of all impacting raindrops because terminal velocity is predictable. The resulting imprinted ash substrates were lithified using hair spray and low-viscosity liquid urethane plastic. The dimensions were measured using the same laser scanner as that used for the latex peels. Each imprinted tray captured a dozen imprints originating from raindrops of the same mass, from which a mean and standard deviation were obtained.

We followed published methods¹⁹ to predict theoretically from first principles how raindrop terminal velocity changes with air density, and thus how dimensionless momentum changes with air density. Given the measurement of the largest Ventersdorp imprint, we obtained the corresponding dimensionless momentum of the impacting drop using our experimental relationship. By assuming the dimension of the raindrop responsible for the largest imprint (bounded by the maximum diameter of 6.8 mm), we quantified atmospheric density (Supplementary Information).

Received 8 July 2011; accepted 23 January 2012.

Published online 28 March 2012.

1. Goldblatt, C. *et al.* Nitrogen-enhanced greenhouse warming on early Earth. *Nature Geosci.* **2**, 891–896 (2009).
2. Kasting, J. Theoretical constraints on oxygen and carbon dioxide concentrations in the Precambrian atmosphere. *Precamb. Res.* **34**, 205–229 (1987).
3. Ohmoto, H., Watanabe, Y. & Kumazawa, K. Evidence from massive siderite beds for a CO₂-rich atmosphere before 1.8 billion years ago. *Nature* **429**, 395–399 (2004).
4. Claire, M., Catling, D. & Zahnle, K. Biogeochemical modelling of the rise in atmospheric oxygen. *Geobiology* **4**, 239–269 (2006).
5. Domagal-Goldman, S., Kasting, J., Johnston, D. & Farquhar, J. Organic haze, glaciations and multiple sulfur isotopes in the Mid-Archaean Era. *Earth Planet. Sci. Lett.* **269**, 29–40 (2008).
6. Haqq-Misra, J., Domagal-Goldman, S., Kasting, P. & Kasting, J. A revised, hazy methane greenhouse for the Archaean Earth. *Astrobiology* **8**, 1127–1137 (2008).
7. Driese, S. G. *et al.* Neoproterozoic paleoweathering of tonalite and metabasalt: implications for reconstructions of 2.69 Ga early terrestrial ecosystems and paleoatmospheric chemistry. *Precamb. Res.* **189**, 1–17 (2011).

8. Kasting, J. & Howard, M. Atmospheric composition and climate on the early Earth. *Phil. Trans. R. Soc. B* **361**, 1733–1742 (2006).
9. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
10. Holland, H. D. The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361**, 903–915 (2006).
11. Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
12. Garvin, J. *et al.* Isotopic evidence for an aerobic nitrogen cycle in the latest Archean. *Science* **323**, 1045–1048 (2009).
13. Knauth, L. & Lowe, D. High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geol. Soc. Am. Bull.* **115**, 566–580 (2003).
14. Lyell, C. On fossil rain-marks of the Recent, Triassic, and Carboniferous periods. *Q. J. R. Geol. Soc.* **7**, 238–247 (1851).
15. Spilhaus, A. Raindrop size, shape and falling speed. *J. Atmos. Sci.* **5**, 108–110 (1948).
16. Magono, C. On the shape of water drops falling in stagnant air. *J. Meteorol.* **11**, 77–79 (1954).
17. Matthews, J. & Mason, B. Electrification produced by the rupture of large water drops in an electric field. *Q. J. R. Meteorol. Soc.* **90**, 275–286 (1964).
18. Lorenz, R. The life, death and afterlife of a raindrop on Titan. *Planet. Space Sci.* **41**, 647–655 (1993).
19. Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* (Kluwer Academic, 1997).
20. Willis, P. & Tattelman, P. Drop-size distribution associated with intense rainfall. *J. Appl. Meteorol.* **28**, 3–15 (1989).
21. Dodd, K. On the disintegration of water drops in an air stream. *J. Fluid Mech.* **9**, 175–182 (1960).
22. Beard, K. V. Velocity and shape of cloud and precipitation drops aloft. *J. Atmos. Sci.* **33**, 851–864 (1976).
23. Clift, R., Grace, J. R. & Weber, M. E. *Bubbles, Drops and Particles* Ch. 7 (Academic, 1978).
24. Foote, G. B. & du Toit, P. S. Terminal velocity of raindrops aloft. *J. Appl. Meteorol.* **8**, 249–253 (1969).
25. van der Westhuizen, W., Grobler, N., Looek, J. & Tordiffe, E. Raindrop imprints in the Late Archaean–Early Proterozoic Ventersdorp Supergroup, South Africa. *Sedim. Geol.* **61**, 303–309 (1989).
26. Reineck, H. & Singh, I. S. *Depositional Sedimentary Environments* 61 (Springer, 1980).
27. Huang, C., Bradford, J. M. & Cushman, J. H. A numerical study of raindrop impact phenomena: the elastic deformation case. *Soil Sci. Soc. Am. J.* **47**, 855–861 (1983).
28. Easton, R. Stratigraphy of Kilauea Volcano. In *Volcanism in Hawaii* (eds Decker, R., Wright, T. & Stauffer, P.) 243–260 (US Government Printing Office, US Geol. Surv. Prof. Pap. 1350, 1987).
29. Zachos, J., Dickens, G. & Zeeber, R. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
30. Hren, M., Tice, M. & Chamberlain, C. Oxygen and hydrogen isotope evidence for a temperate climate 3.42 billion years ago. *Nature* **462**, 205–208 (2009).
31. Knauth, L. Temperature and salinity history of the Precambrian ocean: implications for the course of microbial evolution. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **219**, 53–69 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NASA Exobiology/Astrobiology grant NNX08AP56G. We thank W. Van der Westhuizen of the University of the Free State in South Africa, and E. and D. Jackson of Omdraivlei for their assistance when sampling in the field. We also thank E. Stüeken, A. Chen and K. Huntington at the University of Washington for laboratory assistance, and the staff at Metron Corporation for data acquisition. XRF measurements were performed by the Washington State University Geoanalytical Laboratory. Funding and field logistics for the Iceland fieldwork was supported by the Coordination Action for Research Activities on life in Extreme Environments (CAREX), a project supported by the European Commission Seventh Framework Programme. Funding and field logistics for the Hawaiian fieldwork was supported by the University of Washington Department of Earth and Space Sciences, and its Geoclub. D.C.C. was also supported by NASA Exobiology/Astrobiology grant NNX10AQ90G.

Author Contributions D.C.C. conceived the research project and established maximum-raindrop-size terminal velocity dependence on air density, R.B. led the field work in South Africa, collected the latex peels and analysed the grain sizes of the Ventersdorp tuff, J.P.H. found additional geographic and stratigraphic occurrences of raindrop imprints while performing field work and helping collect latex peels, P.M.P. collected experimental data and measured the Eyjafjallajökull and Pahala ash grain sizes, and S.M.S. developed the method of dimensionless momentum, collected the volcanic ash from Hawaii and Iceland, analysed the data, and led the experimental work. S.M.S., R.B. and D.C.C. discussed results and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M.S. (sanjoy.m.som@nasa.gov).

Formation of the ‘Great Unconformity’ as a trigger for the Cambrian explosion

Shanan E. Peters¹ & Robert R. Gaines²

The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations^{2–5}, but the mechanisms responsible for those perturbations remain uncertain^{1,2}. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁶. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the ‘Cambrian explosion’ of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally tilted sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{7,8}, Baltica⁹, Avalonia⁹ and Siberia¹⁰, making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first skeletonized crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution^{1,6}.

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{11,12} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{11,12}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth’s surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth’s crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreface system, often called the ‘wave-base razor’¹³, as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors⁷ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith^{14,15}, and

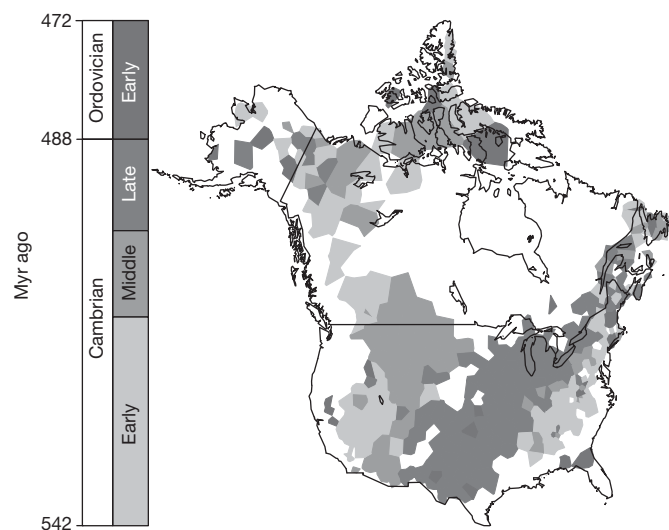


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geoscience, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.

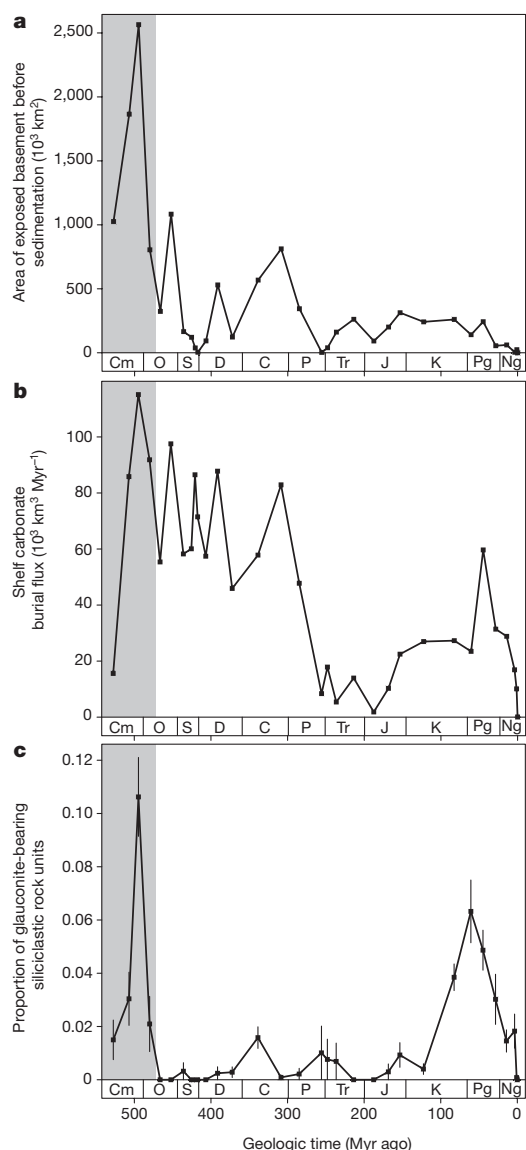


Figure 2 | Phanerozoic sedimentation patterns in North America.

a, Minimum area of crystalline basement rock exposed at surface before burial by sediments in each epoch. **b**, Estimated minimum total carbonate burial flux in each epoch, based on lithostratigraphic units that are composed primarily of carbonate (Fig. 3c). **c**, Proportion of glauconite-rich siliciclastic sedimentary rock units (Fig. 3d). Grey bars span the Cambrian–Early Ordovician Sauk Sequence. Error bars, ± 1 standard error. O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Palaeogene; Ng, Neogene.

because the chemical weathering of basement-derived silicate minerals in soils and regolith is enhanced by physical reworking¹⁵. The sharp stratigraphic contact that typically separates Precambrian basement from the quartz- and feldspar-rich marine sands of the basal, time-transgressive Sauk Sequence (Fig. 3e; see Supplementary Information) demonstrates the efficacy of transgressive shoreface systems in stripping soils and regolith. Shorter-duration eustatic sea level falls, superimposed on the longer-term Sauk transgression, as well as at the terminal Ediacaran², provided an additional mechanism for subaerial exposure, reworking and chemical weathering of basement- and regolith-derived sediments.

Exposure and chemical weathering of silicate minerals derived from continental basement rocks during the formation of the Great Unconformity exerted an important control on seawater chemistry and global biogeochemical cycling in several ways, including the consumption of atmospheric CO_2 and the release of HCO_3^- , Ca^{2+} ,

H_3SiO_4^- , SO_4^{2-} , Cl^- , Na^+ , Mg^{2+} , K^+ , Fe^{3+} and other ions to the oceans^{14,15}. We find sedimentological evidence for an increased reservoir size and a sustained flux of continental weathering products to the ocean within each of the three primary facies belts that onlap Cambrian palaeocontinents¹⁶: the nearshore, generally coarse-grained inner detrital belt, the middle carbonate platform belt, and the outer, fine-grained detrital belt (Fig. 3).

The signal of enhanced continental crustal weathering is perhaps most conspicuously expressed by precipitation of carbonate sediments, which reached a Phanerozoic peak in shelf burial flux during the Sauk transgression (Figs 2b, 3c)^{17,18}. In Laurentia, the large quantity of Cambrian–Early Ordovician carbonates is known as the ‘Great American Bank’¹⁹. Precipitation of carbonates is an important sink for alkalinity that is derived from chemical weathering, and we find that the area of continental basement exposed before sedimentation in each time interval (Fig. 2a) is predictive of changes in carbonate burial rates throughout the Phanerozoic (Fig. 2b; partial linear product-moment correlation between first differences, with effect of varying interval duration held constant, is 0.74, $P = 5 \times 10^{-6}$). These results are consistent with a recent model of the Cambrian carbon cycle, which demonstrated that unusually large absolute rates of carbon throughput are required to explain global carbon isotopic excursions². In the absence of a deep-sea sink for carbonate, which did not become established until the Mesozoic origin of calcareous plankton^{18,20}, Cambrian carbonate burial rates were limited by the rate at which shelf accommodation space was created by sea level rise during the Sauk transgression and by lithospheric subsidence. Thus, it is likely that in the earliest Cambrian ocean there was both a large pre-existing reservoir of chemical weathering products and an enhanced flux originating from the weathering of continental crust.

Additional evidence for an unusually high flux of continental weathering products to the Cambrian ocean, and a corresponding increase in the extent of carbonate precipitation, is provided by petrographic and geochemical data from outer detrital belt mudstones²¹. Although the Phanerozoic rock record is replete with carbonate cemented mudstones, most cements form as a result of microbially mediated remineralization of organic matter and, therefore, have $\delta^{13}\text{C}$ values that are significantly depleted relative to sea water²². Outer detrital belt mudstones of the Sauk Sequence, by contrast, not only contain a significantly greater wt% carbonate than composite shale standards, but they also yield $\delta^{13}\text{C}$ values that are consistent with direct precipitation from sea water (see Supplementary Information). Petrographic analysis reveals that carbonate cements have textures indicative of rapid, displacive growth at bed tops, near the sediment–water interface²¹ (Fig. 3b) and that detrital micrite is rare or absent. Furthermore, many outer detrital belt mudstones contain carbonate concretions and/or mud mounds with $\delta^{13}\text{C}$ values indicative of precipitation from seawater-derived bicarbonate²³ (see Supplementary Information).

The signal of enhanced weathering of continental crust during the Sauk transgression is evidenced in the inner detrital belt by a peak in the frequency of glauconite-rich siliciclastic rock units (Figs 2c, 3d). Glauconite, $(\text{K},\text{Na})(\text{Fe}^{3+},\text{Al},\text{Mg})_2(\text{Si},\text{Al})_4\text{O}_{10}(\text{OH})_2$, is a phyllosilicate mineral that in modern marine environments forms at depths of >50 m and in regions where sedimentation rates are low and prolonged periods of pore-water exchange with seawater occur²⁴. However, it has been noted that the distribution and abundance of glauconite in Cambrian marine sediments (Fig. 2c) required rapid authigenesis under chemical conditions that are different from those found on modern marine shelves^{9,25}. We attribute this anomaly in glauconite formation to an unusually large flux of continental chemical weathering products, particularly K^+ , Fe^{3+} and H_3SiO_4^- , during the formation of the Great Unconformity. Restricted circulation and low net rates of sedimentation within the inner detrital belt may have also contributed to the unusual distribution and abundance of glauconite.

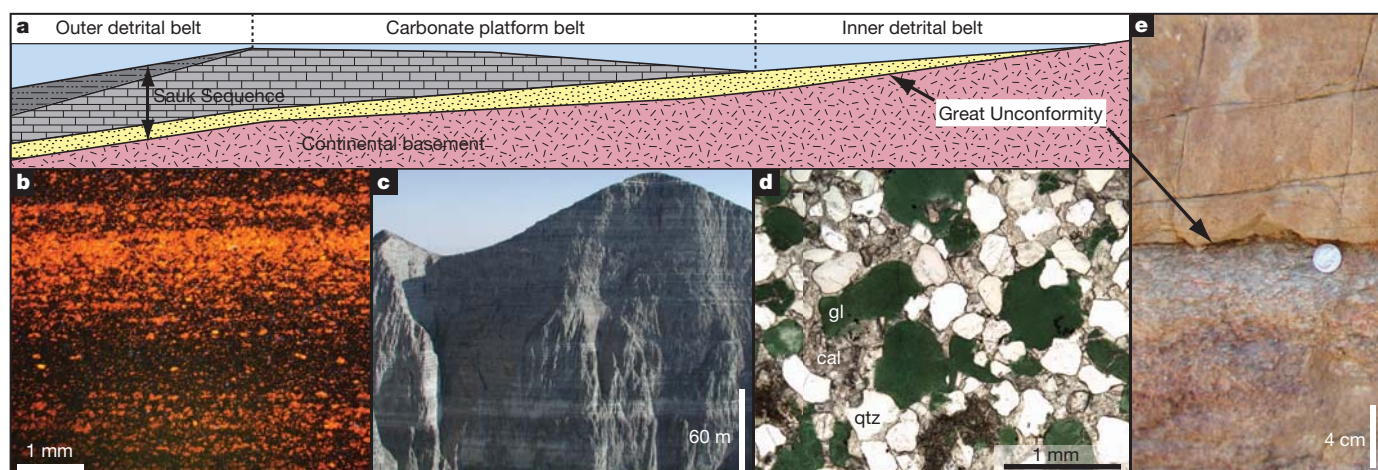


Figure 3 | Middle to Late Cambrian palaeoenvironments and sedimentology. **a**, Generalized marine shelf profile¹⁶. **b**, Cathodoluminescence photomicrograph of the Kaili Formation, an outer detrital belt mudstone from south China. Lamination tops are defined by authigenic carbonate cements that glow orange under cathodoluminescence; clay-rich lamina bases are dark. **c**, Exposure of Notch Peak Formation in Utah. **d**, Plane-polarized light

photomicrograph of a carbonate-cemented, glauconite-bearing sandstone from the Au Train Formation, inner detrital belt, northern Michigan. **e**, Great Unconformity in Wind River Canyon, Wyoming, with ~510-Myr-old marine Flathead Sandstone on 2,900-Myr-old granitic basement. Coin is US dime (17.9 mm diameter). gl, glauconite; cal, calcite; qtz, quartz.

Other records of seawater chemistry provide supporting evidence for changes in tectonic activity and enhanced continental weathering during the formation of the Great Unconformity. For example, the average $^{87}\text{Sr}/^{86}\text{Sr}$ of sea water increased during the Neoproterozoic (Fig. 4). This long-term signal is attributable to the increasing concentration of ^{87}Sr in continental crust due to the decay of ^{87}Rb , and to the long-term erosion and progressive exposure of ^{87}Sr -rich granitic rocks required to form the Great Unconformity. During the Cambrian, $^{87}\text{Sr}/^{86}\text{Sr}$ increased more rapidly to achieve a 900 Myr maximum (that is, a maximum over the past 900 Myr) near the end of the Sauk transgression², a signal we attribute to enhanced weathering of continental crust during the Sauk transgression. Although Phanerozoic $^{87}\text{Sr}/^{86}\text{Sr}$ values exhibit many shorter-term oscillations, average $^{87}\text{Sr}/^{86}\text{Sr}$ values declined after the Late Cambrian peak for most of the Phanerozoic (Fig. 4). We hypothesize that this reversal in the long-term trend in seawater $^{87}\text{Sr}/^{86}\text{Sr}$ was driven by progressive Phanerozoic sedimentary re-covering of ^{87}Sr -rich continental crust and by increased contributions of ^{86}Sr from the hydrothermal alteration of mid-ocean-ridge basalt. The 900 Myr Cambrian peak in $^{87}\text{Sr}/^{86}\text{Sr}$ was not approached again until the recent, possibly indicating enhanced removal of

Phanerozoic sedimentary cover during Cenozoic orogenesis and global climate change¹⁵.

Congruent evidence for enhanced weathering of continental crust during the Sauk transgression is provided by ϵNd , an isotopic tracer with a short oceanic residence time that is sensitive to the mean age of continental crust undergoing chemical weathering. Average seawater ϵNd declined through the Cambrian to reach a long-term minimum at the peak of the Sauk transgression²⁶ (Fig. 4), reflecting increasingly important input from ancient continental basement rocks as the Sauk transgression progressed from relatively young continental margins into generally older cratonic interiors.

Perhaps the strongest evidence for an increased reservoir and flux of continental chemical weathering products to the global ocean is provided by direct measurements of $[\text{Ca}^{2+}]$ in evaporite fluid inclusions. Concentrations of Ca^{2+} in sea water rose precipitously from the Neoproterozoic to reach a Phanerozoic high in the Cambrian^{27–29}. We attribute much of this approximately threefold increase in seawater $[\text{Ca}^{2+}]$ to enhanced chemical weathering of continental crust during the Sauk transgression. Elevated seawater $[\text{Ca}^{2+}]$ has been proposed as a mechanism for the origin of biocalcification by promoting intracellular

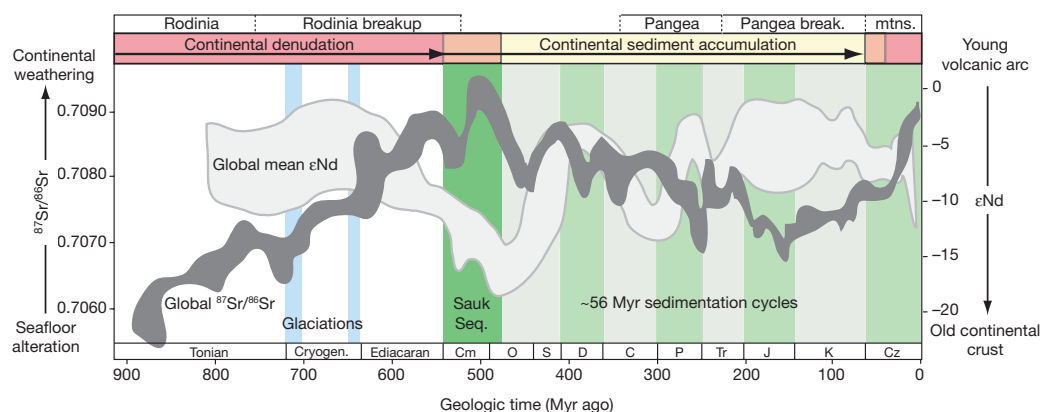


Figure 4 | Summary of major tectonic, geochemical and sedimentary patterns over the past 900 Myr. The shift from widespread continental denudation to widespread sedimentation on the continents defines the Great Unconformity. $^{87}\text{Sr}/^{86}\text{Sr}$ estimates (dark grey envelope) from ref. 2 and references therein; ϵNd estimates (light grey envelope) from ref. 26. Vertical blue bars indicate times of major global glaciations; vertical green bars identify continent-

scale sedimentary sequences¹², including the Cambrian–Early Ordovician Sauk Sequence (dark green bar). Approximate times of widespread continental denudation are identified by horizontal red bars at top; time of widespread sediment accumulation on continents is identified by partially overlapping light yellow bar. Periods of major supercontinents (Rodinia, Pangea) and their breakup are identified, including recent mountain building (mntns.).

Ca^{2+} concentrations that were toxic to some animals^{28,29}. Here we provide a mechanism for this increase in seawater $[\text{Ca}^{2+}]$, as well as other ions involved in biomineralization.

The palaeogeomorphic surface represented by the Great Unconformity is unique in at least the past 900 Myr of Earth history, but similar episodes of widespread continental denudation followed by large-scale marine transgression and sedimentation may have occurred before 900 Myr ago. However, the advent of biomineralization and the Cambrian explosion required the presence of animals, which evolved in Cryogenian to Ediacaran time^{1,2,30}. Our results therefore offer a new hypothesis for the timing and origin of biomineralization and the Cambrian explosion, both of which lag by tens of millions of years the initial origin of bilaterian animals³⁰. Specifically, we propose that biomineralization in multiple clades occurred in response to the changes in global ocean chemistry^{28,29} that were promoted by the formation of the Great Unconformity and by the expansion of shallow marine environments that it records. Although Darwin and other palaeontologists have regarded the resultant widespread hiatus in the rock record as a failure of preservation, the formation of this prominent gap may have actually been an environmental trigger for biomineralization, thereby promoting the Cambrian explosion of marine animals³⁰. Determining the geodynamic causes of extensive Neoproterozoic continental denudation followed by Phanerozoic sedimentation, and linking those dynamics to the timing and spatial distribution of marine transgression and biogeochemical change, is now a challenge for geoscience.

METHODS SUMMARY

Stratigraphic data derive from Macrostrat (<http://macrostrat.org>), which consists of 21,521 rock units distributed among 830 geographic regions in North America³⁰. Geological ages for the top and bottom of each Phanerozoic rock unit are referenced to one of approximately 82 stage-level time intervals. For this analysis, data were binned into 32 Phanerozoic epochs (median duration 15.6 Myr; see Supplementary Information).

Macrostrat is comprehensive in the sense that all rock units in each region, including those in the subsurface, are included. Rock units have an explicitly identified thickness range and dominant lithology in each region. Many units also have secondary lithologies and, when prominent and characteristic of the unit, additional data on sedimentology, such as mineralogy and sedimentary structures.

To estimate the total number of sedimentary rock packages that directly overlie crystalline basement, all of the rock units older than the base of each Phanerozoic sediment package were tabulated in each region. If there were no older sedimentary rocks at a location, then that package was identified as resting on basement. This approach is conservative with respect to the Great Unconformity because a large fraction of Precambrian sedimentary rock is structurally deformed and laterally adjacent to basement rocks. Thus, the Great Unconformity often cuts across all older rock units at each location⁶.

Carbon and oxygen isotopic compositions of 220 outer detrital belt mudstone samples from North America and China were obtained by dissolving whole rock powders in phosphoric acid and analysing evolved CO_2 at the stable isotope laboratories at the University of California Davis and Riverside. Standard errors for all measurements were $\leq 0.2\text{‰}$. Whole rock geochemistry measurements from 82 outer detrital belt mudstone samples in North America and China were obtained from ultrasonically cleaned, twice-fused polished glass beads using a Panalytical Axios XRF instrument at Pomona College (see Supplementary Information).

Received 29 June 2011; accepted 16 February 2012.

1. Marshall, C. R. Explaining the Cambrian “Explosion” of animals. *Annu. Rev. Earth Planet. Sci.* **34**, 355–384 (2006).
2. Maloof, A. C. *et al.* The earliest Cambrian record of animals and ocean geochemical change. *Geol. Soc. Am. Bull.* **122**, 1731–1774 (2010).
3. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* **315**, 92–95 (2007).
4. Dahl, T. W. *et al.* Devonian rise in atmospheric oxygen correlated to radiations of terrestrial plants and large predatory fish. *Proc. Natl Acad. Sci. USA* **107**, 17911–17915 (2010).
5. Gill, B. C. *et al.* Geochemical evidence for widespread euxinia in the later Cambrian ocean. *Nature* **469**, 80–83 (2011).

6. Yochelson, E. L. The Lipalian interval: a forgotten, novel concept in the geologic column. *Earth Sci. Hist.* **25**, 251–269 (2006).
7. Avigad, D. *et al.* Mass-production of Cambro-Ordovician quartz-rich sandstone as a consequence of chemical weathering of Pan-African terranes: environmental implications. *Earth Planet. Sci. Lett.* **240**, 818–826 (2005).
8. Laird, M. G. in *Geological Evolution of Antarctica* (eds Thomson, M. R. A., Crame, J. A. & Thomson, J. W.) 177–186 (Cambridge Univ. Press, 1991).
9. Brasier, M. D. The Lower Cambrian transgression and glauconite-phosphate facies in western Europe. *J. Geol. Soc. Lond.* **137**, 695–703 (1980).
10. Sears, J. W. & Price, R. A. Tightening the Siberian connection to western Laurentia. *Geol. Soc. Am. Bull.* **115**, 943–953 (2003).
11. Sloss, L. L. Sequences in the cratonic interior of North America. *Geol. Soc. Am. Bull.* **74**, 93–114 (1963).
12. Meyers, S. R. & Peters, S. E. A 56 million year rhythm in North American sedimentation during the Phanerozoic. *Earth Planet. Sci. Lett.* **303**, 174–180 (2011).
13. Cross, T. A. & Lessenger, M. A. in *Innovative Applications of Petroleum Technology in the Rocky Mountain Region* (eds Coalson, E. B., Osmond, J. C. & Williams, E. T.) 183–203 (Rocky Mountain Association of Geologists, Denver, 1997).
14. Mortatti, J. & Probst, J. Silicate rock weathering and atmospheric/soil CO_2 uptake in the Amazon basin estimated from river water geochemistry: seasonal and spatial variations. *Chem. Geol.* **197**, 177–196 (2003).
15. Millot, R., Gaillardet, J., Dupré, B. & Allegre, C. J. The global control of silicate weathering rates and the coupling with physical erosion: new insights from rivers of the Canadian Shield. *Earth Planet. Sci. Lett.* **196**, 83–98 (2002).
16. Robison, R. A. in *Guidebook to the Geology of East Central Nevada* (eds Boettcher, J. W. Jr & Sloan, W. W.) 43–52 (Intermountain Association of Petroleum Geologists, Salt Lake City, 1960).
17. Ronov, A. B., Khain, V. E., Balukhovskiy, A. N. & Seslavinsky, K. B. Quantitative analysis of Phanerozoic sedimentation. *Sedim. Geol.* **25**, 311–325 (1980).
18. Walker, L. J., Wilkinson, B. H. & Ivany, L. C. Continental drift and Phanerozoic carbonate accumulation in shallow-shelf and deep-marine settings. *J. Geol.* **110**, 75–87 (2002).
19. Ginsburg, R. N. Actualistic depositional models for the Great American Bank (Cambro-Ordovician). In *Eleventh International Congress on Sedimentology, Abstracts of Papers 114* (International Association of Sedimentologists/McMaster University, 1982).
20. Ridgwell, A. J., Kennedy, M. J. & Caldeira, K. Carbonate deposition, climate stability, and Neoproterozoic ice ages. *Science* **302**, 859–862 (2003).
21. Gaines, R. R. *et al.* Mechanism for Burgess Shale-type preservation. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1111784109> (published online, 5 March 2012).
22. Coleman, M. L. Geochemistry of diagenetic non-silicate minerals: kinetic considerations. *Phil. Trans. R. Soc. Lond. A* **315**, 39–56 (1985).
23. Elrick, M. & Snider, A. S. Deep-water stratigraphic cyclicity and carbonate mud mound development in the Middle Cambrian Marjum Formation, House Range, Utah, U.S.A. *Sedimentology* **49**, 1021–1047 (2002).
24. Odin, G. S. & Matter, A. De glauconiarum origine. *Sedimentology* **28**, 611–641 (1981).
25. Chafetz, H. S. & Reid, A. Syndepositional shallow-water precipitation of glauconite minerals. *Sedim. Geol.* **136**, 29–42 (2000).
26. Keto, L. S. & Jacobson, S. B. Nd isotopic variations of Phanerozoic palaeoceans. *Earth Planet. Sci. Lett.* **90**, 395–410 (1988).
27. Lowenstein, T. K. *et al.* Oscillations in Phanerozoic seawater chemistry: evidence from fluid inclusions. *Science* **294**, 1086–1088 (2001).
28. Brennan, S. T., Lowenstein, T. K. & Horita, J. Seawater chemistry and the advent of biocalcification. *Geology* **32**, 473–476 (2004).
29. Petrychenko, O. Y., Peryt, T. M. & Chechei, E. I. Early Cambrian seawater chemistry from fluid inclusions in halite from Siberian evaporites. *Chem. Geol.* **219**, 149–161 (2005).
30. Erwin, D. H. *et al.* The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091–1097 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Canfield, P. Cohen, W. Fischer, S. Finnegan, N. Heim, A. Carroll and R. Dott for discussion, and N. Butterfield, M. Foote, E. Hammarlund, P. Myrow, B. Wilkinson, R. Wood, S. Holland for feedback on ideas. Fieldwork and analysis was aided by P. Burke, J. B. Caron, L. Curtin, F. Dwan, Z. Feng, P. Fenton, L. Finley-Blasi, X. Hou, J. Lackey, C. Qi, J. Peng, J. Tian, J. Vorhies, Y. Yang, X. Zhang and Y. Zhao. Work was supported by NSF EAR-0819931 (to S.E.P.) and EAR-1046233 and DUE-0942447 (to R.R.G.).

Author Contributions S.E.P. contributed Macrostrat-derived data, R.R.G. contributed sample-derived data. Both authors contributed to the development of ideas and writing.

Author Information Data for aspects of this analysis derive from Macrostrat (<http://macrostrat.org>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.E.P. (peters@geology.wisc.edu).

Clusters of iron-rich cells in the upper beak of pigeons are macrophages not magnetosensitive neurons

Christoph Daniel Treiber¹, Marion Claudia Salzer¹, Johannes Riegler², Nathaniel Edelman¹, Cristina Sugar¹, Martin Breuss¹, Paul Pichler¹, Herve Cadiou³, Martin Saunders⁴, Mark Lythgoe², Jeremy Shaw⁴ & David Anthony Keays¹

Understanding the molecular and cellular mechanisms that mediate magnetosensation in vertebrates is a formidable scientific problem^{1,2}. One hypothesis is that magnetic information is transduced into neuronal impulses by using a magnetite-based magnetoreceptor^{3,4}. Previous studies claim to have identified a magnetic sense system in the pigeon, common to avian species, which consists of magnetite-containing trigeminal afferents located at six specific loci in the rostral subepidermis of the beak^{5–8}. These studies have been widely accepted in the field and heavily relied upon by both behavioural biologists and physicists^{9–11}. Here we show that clusters of iron-rich cells in the rostro-medial upper beak of the pigeon *Columbia livia* are macrophages, not magnetosensitive neurons. Our systematic characterization of the pigeon upper beak identified iron-rich cells in the stratum laxum of the subepidermis, the basal region of the respiratory epithelium and the apex of feather follicles. Using a three-dimensional blueprint of the pigeon beak created by magnetic resonance imaging and computed tomography, we mapped the location of iron-rich cells, revealing unexpected variation in their distribution and number—an observation that is inconsistent with a role in magnetic sensation. Ultrastructure analysis of these cells, which are not unique to the beak, showed that their subcellular architecture includes ferritin-like granules, siderosomes, haemosiderin and filopodia, characteristics of iron-rich macrophages. Our conclusion that these cells are macrophages and not magnetosensitive neurons is supported by immunohistological studies showing co-localization with the antigen-presenting molecule major histocompatibility complex class II. Our work necessitates a renewed search for the true magnetite-dependent magnetoreceptor in birds.

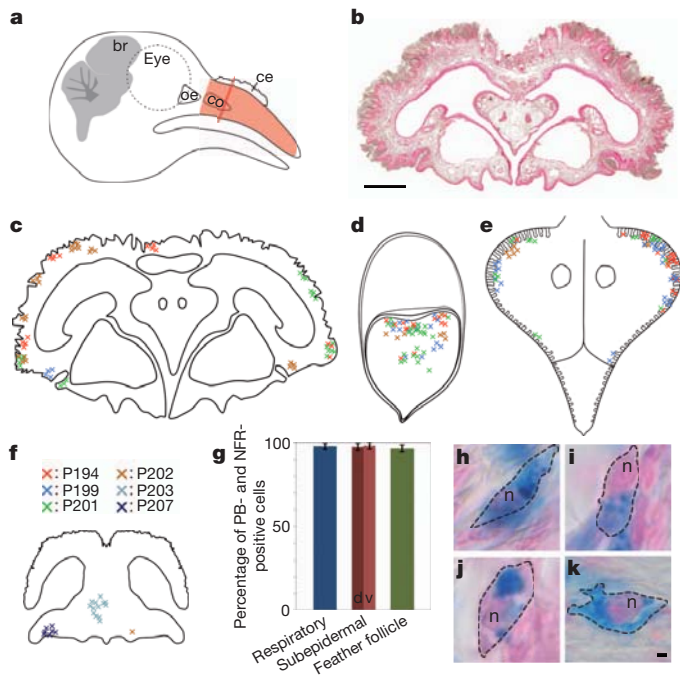
Each year millions of birds complete lengthy journeys guided by the Earth's magnetic field. Current evidence indicates that the detection of magnetic fields is mediated by an inclination-sensitive light-dependent compass that resides in the retina^{12,13}, and an intensity-sensitive apparatus that is believed to provide information about the magnetic map, is associated with the trigeminal nerve, and is thought to rely on biogenic magnetite (Fe₃O₄)¹⁴. The trigeminal nerve was first implicated in magnetoreception in the bobolink *Dolichonyx oryzivorus*, and was suggested to be sensitive to small alterations in magnetic stimuli¹⁵. Subsequent studies revealed that the ophthalmic branch is required for pigeons to perform an intensity-based conditioning task³, and that neurons in the trigeminal brainstem complex of European robins are activated when the birds are subjected to non-uniform magnetic fields¹⁶. These data have led to the proposition that the sensory cells responsible for magnetite-based magnetoreception lie in the upper beak of birds^{3,17}. Previous studies^{3,7} have claimed that clusters of iron-containing neurons in six specific bilateral locations in the rostral dermis of the upper beak of pigeons constitute a magnetic sense system⁷. It has been contended that this system consists of unmyelinated dendrites that contain superparamagnetic spherules surrounded by iron platelets that are composed of magnetite and

maghemite, and that the system is a common sensory apparatus in birds^{5–7}. These assertions have formed the basis for a host of behavioural studies and theoretical calculations that aim to advance the magnetite theory of magnetoreception^{9,10,18–20}.

To investigate this putative magnetic sense system, we undertook a systematic analysis of the prevalence and distribution of all iron-rich cells in the upper beak of the pigeon. We perfused adult pigeons (Nuremberg cohort, $n = 12$), and sectioned the upper beak from the caudal respiratory concha to the tip of the beak in the coronal plane (Fig. 1a, b). We stained serial sections (10 μ m) with Prussian blue (PB) to label ferric iron, and nuclear fast red (NFR) to identify nuclei, followed by counting of all PB-positive cells. We consistently observed PB-positive cells in three specific regions: (1) in the stratum laxum of the dorsal and/or ventral subepidermis; (2) in the buds of feather follicles; and (3) in the basal region of the respiratory epithelium (Fig. 1c–e). We confirmed this pattern of staining in a larger collection of pigeons originating from seven different lofts ($n = 172$). PB-positive cells in all three regions were characterized by the presence of multiple dark blue spherules (0.25–5.0 μ m in size) and/or by light blue cytoplasmic staining with a notable nucleus (Fig. 1g–k and Supplementary Figs 1–3). Subepidermal PB-positive cells in caudal and medial regions were predominantly found in the dorsal subepidermis (Fig. 1c), whereas those PB-positive cells located rostrally were found in the ventral subepidermis lining the inner roof of the beak (Fig. 1f and Supplementary Fig. 1). PB-positive cells in the feather follicle clustered in the apical region of the bud (Fig. 1d and Supplementary Fig. 2), and those in the respiratory epithelium were predominantly found within the lateral edges of the concha (Fig. 1e and Supplementary Fig. 3).

As it is believed that iron-rich cells in the upper pigeon beak are limited to six discrete bilateral anatomical loci⁵, we mapped the distribution of PB-positive cells along the rostro-caudal axis of the beak. To do this accurately we first created a three-dimensional topographic map of the pigeon beak by undertaking high-resolution magnetic resonance imaging (MRI) coupled with micro-computed tomography (micro-CT) scanning, identifying four specific anatomical landmarks (Supplementary Movies 1, 2 and Supplementary Figs 4, 5). After staining serial sections, we counted PB-positive cells and used our landmarks to map the distribution of cells along the rostro-caudal axis ($n = 12$). We found that PB-positive cells in the respiratory epithelium and feather follicles were restricted to caudal regions, whereas those in the subepidermal region were found in clusters along the length of the beak with no apparent bilateralization (Fig. 2a–c and Supplementary Figs 6, 7 and Supplementary Table 1). We found no significant differences in the total number of PB-positive cells between sexes (respiratory epithelium ($P > 0.5$), subepidermis ($P > 0.1$), feather follicle ($P > 0.1$)) (Supplementary Fig. 8), but observed an extremely large variation in the number and distribution of PB-positive cells when comparing birds of the same age and sex. For instance, pigeon 200 had ~200 PB-positive cells in the subepidermis, whereas pigeon

¹Institute of Molecular Pathology, Dr Bohr-Gasse, 1030 Vienna, Austria. ²Centre for Advanced Biomedical Imaging (CABI), Department of Medicine and Institute of Child Health, University College London (UCL), London WC1E 6DD, UK. ³Université de Strasbourg, Institut des Neurosciences Cellulaires et Intégratives (INCI), CNRS UPR 3212, F-67084 Strasbourg, France. ⁴Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, Crawley 6009, Australia.

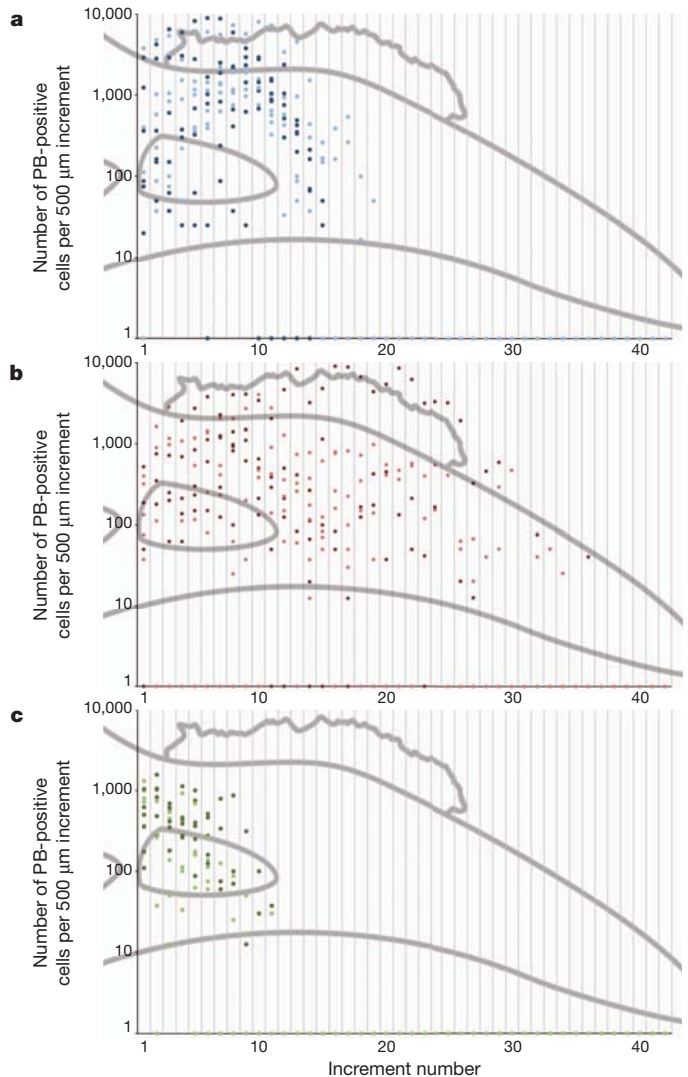


203 had ~108,800 PB-positive cells located in numerous clusters along the length of the beak. Although our serial quantification samples a section every 120 μ m we did not find the six 350- μ m-long bilateral clusters that are claimed to constitute a magnetic sense system⁷. We speculated that our pigeon strain might harbour a large genomic deletion that accounts for the absence of this putative magnetosensitive system. To investigate this we quantified PB-positive cells in pigeons from another loft (Vienna cohort, $n = 6$). Similar to our Nuremberg cohort we did not find six bilateral clusters, and once again observed a large variation in the distribution and number of PB-positive cells (Supplementary Fig. 9 and Supplementary Table 2). This variation is not consistent with a genetically encoded sensory apparatus responsible for magnetosensation.

Next we asked whether PB-positive cells are neurons by triple staining sections with PB, NFR, and one of three different antibodies that label neuronal structures: neurofilament (NF), TUBB3 and MAP1B ($n \geq 5$ birds). In the respiratory epithelium we observed 0.04% co-localization with NF ($n = 1,208$ cells), 0.6% co-localization with TUBB3 ($n = 2,818$ cells), and 0.01% co-localization with MAP1B ($n = 2,213$ cells). In the subepidermis we found no co-localization with NF ($n = 471$ cells) or MAP1B ($n = 803$ cells), and only 0.06% co-localization with TUBB3 ($n = 1,309$ cells). Finally, in the feather follicle we found no co-localization with NF ($n = 286$ cells) or MAP1B ($n = 295$ cells), and only 0.24% co-localization with TUBB3 ($n = 407$ cells) (Supplementary Figs 10, 11). The simplest explanation for the

very small amount of apparent co-localization we observed is that two cells, one PB positive and the other positive for a neuronal marker, lie in the same vertical plane, and because of the nature of the chemical stain used cannot be distinguished from one another. Taken together, our results strongly suggest that the clusters of PB-positive cells in the beak of the pigeon are not neurons.

To ascertain the true identity of the PB-positive cells, we undertook an analysis of their ultrastructure using transmission electron microscopy (TEM) ($n = 3$ birds) (Supplementary Fig. 12). We observed ferritin-like granules (6–9 nm) throughout the cytoplasm of PB-positive cells from all regions and in some instances haemosiderin masses and/or membrane-bound electron-dense organelles known as siderosomes^{21,22} (~300 nm) (Fig. 3a–f and Supplementary Fig. 13). Energy-filtered transmission electron microscopy (EFTEM) confirmed that each electron-dense granule was composed of iron (Supplementary



very small amount of apparent co-localization we observed is that two cells, one PB positive and the other positive for a neuronal marker, lie in the same vertical plane, and because of the nature of the chemical stain used cannot be distinguished from one another. Taken together, our results strongly suggest that the clusters of PB-positive cells in the beak of the pigeon are not neurons.

To ascertain the true identity of the PB-positive cells, we undertook an analysis of their ultrastructure using transmission electron microscopy (TEM) ($n = 3$ birds) (Supplementary Fig. 12). We observed ferritin-like granules (6–9 nm) throughout the cytoplasm of PB-positive cells from all regions and in some instances haemosiderin masses and/or membrane-bound electron-dense organelles known as siderosomes^{21,22} (~300 nm) (Fig. 3a–f and Supplementary Fig. 13). Energy-filtered transmission electron microscopy (EFTEM) confirmed that each electron-dense granule was composed of iron (Supplementary

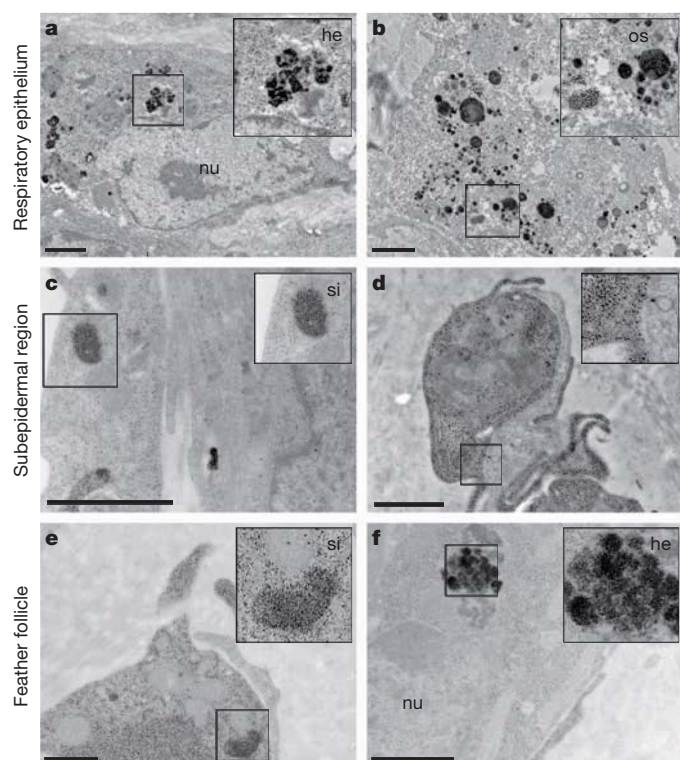


Figure 3 | Ultrastructure of PB-positive cells. a–f, Representative electron micrographs of two different PB-positive cells in the respiratory epithelium (a, b), subepidermal region (c, d) and the feather follicle (e, f; $n = 3$ birds). Cells in all regions were found to contain ferritin-like granules (6–9 nm in diameter) that are present throughout the cytoplasm. In addition we observed haemosiderin (he) clumps and/or membrane bound siderosomes (si). Osmophilic lipid droplets (os) are visible in cells in the respiratory epithelium in b. Cells in the subepidermal region and feather follicle were noted for their slender cytoplasmic projections resembling filopodia, that are seen to engulf a cell in d. Cells in all regions were nucleated (nu). See also Supplementary Fig. 13. Scale bars, 1 μ m.

Fig. 14). Selected area electron diffraction (SAED) failed to identify any cellular structures that contained magnetite, but showed that haemosiderin masses in the feather follicle consist of a goethite-like material ($n = 3$ birds), whereas siderosomes in the respiratory epithelium are comprised of ferrihydrite ($n = 2$ birds) (Supplementary Fig. 15 and Supplementary Table 3). On a cellular level, PB-positive cells in the respiratory epithelium are characterized by the presence of osmophilic lipid vacuoles (Fig. 3b), whereas those cells originating from the feather follicle and subepidermal region had notable dendritic extensions that resembled filopodia (Supplementary Fig. 13d–i, k). In some instances these cytoplasmic tentacles appeared to engulf neighbouring cells, suggesting to us that the PB cells may be phagocytic macrophages (Fig. 3d).

Macrophages are known to reside in the spleen, dermis and respiratory mucosa of multiple species, and to have a vital role in host defence and iron homeostasis²³. Iron accumulates within macrophages during the catabolism of haemoglobin and is stored as ferritin²⁴. In one class of macrophages known as siderophages, ferritin accumulates in membrane-bound siderosomes, which can then undergo proteolytic processing forming haemosiderin. This accumulation of iron renders these cells PB positive^{25,26}. To ascertain whether the PB-positive cells in the upper beak of the pigeon were siderophages, we stained cryosections with sera against major histocompatibility complex class II (MHC II), which labels antigen-presenting cells including macrophages, alongside positive and negative controls ($n \geq 4$ birds) (Fig. 4a–c and Supplementary Fig. 11)²⁷. We observed MHC II co-localization with 98.8% of PB-positive cells in the respiratory epithelium ($n = 104$ cells),

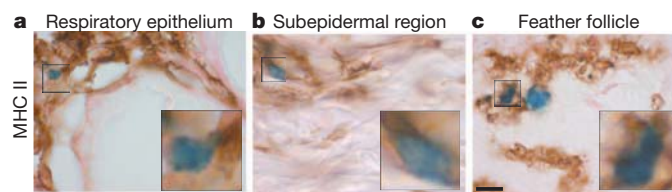


Figure 4 | MHC II immunohistochemistry on PB-positive cells.

a–c, Representative images of coronal sections triple stained with PB, NFR, and sera against the antigen-presenting marker MHC II ($n \geq 4$ birds). MHC II staining was predominant on the surface of cells and found to co-localize with $\geq 94\%$ of PB-positive cells in all regions. Controls are shown in Supplementary Fig. 11. Scale bar, 10 μ m.

95% of PB-positive cells in the subepidermis ($n = 92$ cells), and 94.4% in the feather follicle ($n = 205$ cells). Taken together with our anatomical mapping, subcellular data and neuronal staining, we conclude that clusters of PB-positive cells in the upper beak of the pigeon *C. livia* are macrophages, not magnetosensitive neurons.

As macrophages are not unique to the upper beak, our finding predicts that PB-positive cells should be found throughout the pigeon. To test this we stained skin samples from the back, abdomen, neck, scalp, wing and lower beak of the pigeon ($n \geq 3$ birds). This revealed widespread PB-positive staining in the subepidermis and feather follicle, which was indistinguishable from that observed in the upper beak (Supplementary Fig. 16). Our conclusion further predicts the infiltration of PB-positive macrophages in response to tissue damage or host invasion. We observed such a response in the beak of one of our pigeons (P199), where a large inflammatory lesion with a necrotic centre was surrounded by lymphoplasmacytic cells, including $\sim 80,000$ PB-positive cells that were again characterized by constellations of blue spherules and/or by light blue cytoplasmic staining (Supplementary Fig. 17).

Although we cannot exclude the possibility that a small number of sparsely distributed magnetoreceptors reside at an unknown location in the upper beak of pigeons, this study finds no evidence to support the existence of a subepidermal magnetic sense system that consists of iron-containing dendrites at six specific bilateral loci. This conclusion, which is supported by a critical analysis of the elemental composition of PB-positive cells in the subepidermis²⁸, has several important implications. First, it requires a re-evaluation of behavioural studies that have purported to impair the function of a magnetite-based receptor in the subepidermis of the upper beak and the conclusions that these studies reached^{9,20}. Second, it necessitates a re-assessment as to whether superparamagnetic magnetite has the necessary physical and magnetic properties to act as a magnetosensor in a living system^{10,11,19}. Third, our work reveals that the sensory cells that are responsible for trigeminally mediated magnetic sensation in birds remain undiscovered. These enigmatic cells may reside in the olfactory epithelium, a sensory structure that has been implicated in magnetoreception in the rainbow trout²⁹.

METHODS SUMMARY

Histological studies. We perfused adult pigeons with 4% phosphate-buffered paraformaldehyde (PFA, pH 7.4), and dissected the tissue with ceramic-coated tools. We embedded the tissue in paraffin and prepared 10- μ m coronal sections. We stained sections for 20 min in 5% potassium hexacyanoferrate with 10% HCl, followed by a series of washes and a 2 min exposure to NFR. For immunohistochemistry we incubated the sections with the primary antibody for 18 h in 0.1% Triton PBS with 2–4% milk (pH 7.4), before detection using standard methods, followed by PB staining.

Ultrastructure studies. Adult pigeons were perfused with 2.5% glutaraldehyde supplemented with 2% PFA in PBS (pH 7.4). Tissue was dissected with ceramic-coated tools, incubated for 1 h in phosphate-buffered 2% osmium (pH 7.4), dehydrated and embedded in epoxy resin. We prepared alternative semi-thin (2 μ m) and ultra-thin (70 nm) sections. Semi-thin sections were stained with PB, and the ultra-thin sections were used for TEM, EFTEM and SAED.

Imaging studies. Adult pigeons were perfused with 4% PFA. Following post-fixation and mounting, MRI imaging was performed on a horizontal bore 9.4 T DirectDrive VNMRs system (Agilent Technologies), and CT on a Nucline Nano SPECT/CT imaging system (Mediso).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 December 2011; accepted 19 March 2012.

Published online 15 April 2012.

- Mouritsen, H. & Ritz, T. Magnetoreception and its use in bird navigation. *Curr. Opin. Neurobiol.* **15**, 406–414 (2005).
- Johnsen, S. & Lohmann, K. J. The physics and neurobiology of magnetoreception. *Nature Rev. Neurosci.* **6**, 703–712 (2005).
- Mora, C. V., Davison, M., Wild, J. M. & Walker, M. M. Magnetoreception and its trigeminal mediation in the homing pigeon. *Nature* **432**, 508–511 (2004).
- Kirschvink, J. L., Walker, M. M. & Diebel, C. E. Magnetite-based magnetoreception. *Curr. Opin. Neurobiol.* **11**, 462–467 (2001).
- Fleissner, G. *et al.* Ultrastructural analysis of a putative magnetoreceptor in the beak of homing pigeons. *J. Comp. Neurol.* **458**, 350–360 (2003).
- Falkenberg, G. *et al.* Avian magnetoreception: elaborate iron mineral containing dendrites in the upper beak seem to be a common feature of birds. *PLoS ONE* **5**, e9231 (2010).
- Fleissner, G., Stahl, B., Thalau, P., Falkenberg, G. & Fleissner, G. A novel concept of Fe-mineral-based magnetoreception: histological and physicochemical data from the upper beak of homing pigeons. *Naturwissenschaften* **94**, 631–642 (2007).
- Hanzlik, M. *et al.* Superparamagnetic magnetite in the upper beak tissue of homing pigeons. *Biometals* **13**, 325–331 (2000).
- Wiltchko, R., Schiffner, I., Fuhrmann, P. & Wiltchko, W. The role of the magnetite-based receptors in the beak in pigeon homing. *Curr. Biol.* **20**, 1534–1538 (2010).
- Solov'yov, I. A. & Greiner, W. Theoretical analysis of an iron mineral-based magnetoreceptor model in birds. *Biophys. J.* **93**, 1493–1509 (2007).
- Davila, A. F., Winklhofer, M., Shcherbakov, V. P. & Petersen, N. Magnetic pulse affects a putative magnetoreceptor mechanism. *Biophys. J.* **89**, 56–63 (2005).
- Zapka, M. *et al.* Visual but not trigeminal mediation of magnetic compass information in a migratory bird. *Nature* **461**, 1274–1277 (2009).
- Ritz, T., Thalau, P., Phillips, J. B., Wiltchko, R. & Wiltchko, W. Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature* **429**, 177–180 (2004).
- Cadiou, H. & McNaughton, P. A. Avian magnetite-based magnetoreception: a physiologist's perspective. *J. R. Soc. Interface* **7** (suppl. 2), S193–S205 (2010).
- Beason, R. C. & Semm, P. Magnetic responses of the trigeminal nerve system of the bobolink (*Dolichonyx oryzivorus*). *Neurosci. Lett.* **80**, 229–234 (1987).
- Heyers, D., Zapka, M., Hoffmeister, M., Wild, J. M. & Mouritsen, H. Magnetic field changes activate the trigeminal brainstem complex in a migratory bird. *Proc. Natl Acad. Sci. USA* **107**, 9394–9399 (2010).
- Williams, M. N. & Wild, J. M. Trigeminally innervated iron-containing structures in the beak of homing pigeons, and other birds. *Brain Res.* **889**, 243–246 (2001).
- Wiltchko, W., Munro, U., Ford, H. & Wiltchko, R. Avian orientation: the pulse effect is mediated by the magnetite receptors in the upper beak. *Proc. Biol. Sci.* **276**, 2227–2232 (2009).
- Solov'yov, I. A. & Greiner, W. Micromagnetic insight into a magnetoreceptor in birds: existence of magnetic field amplifiers in the beak. *Phys. Rev. E* **80**, 041919 (2009).
- Stapput, K., Thalau, P., Wiltchko, R. & Wiltchko, W. Orientation of birds in total darkness. *Curr. Biol.* **18**, 602–606 (2008).
- Iancu, T. C. Ferritin and hemosiderin in pathological tissues. *Electron Microsc. Rev.* **5**, 209–229 (1992).
- Richter, G. W. The iron-loaded cell—the cytopathology of iron storage. A review. *Am. J. Pathol.* **91**, 362–404 (1978).
- Wang, J. & Pantopoulos, K. Regulation of cellular iron metabolism. *Biochem. J.* **434**, 365–381 (2011).
- Mebius, R. E. & Kraal, G. Structure and function of the spleen. *Nature Rev. Immunol.* **5**, 606–616 (2005).
- Meguro, R. *et al.* The presence of ferric and ferrous iron in the nonheme iron store of resident macrophages in different tissues and organs: histochemical demonstrations by the perfusion-Perls and -Turnbull methods in the rat. *Arch. Histol. Cytol.* **68**, 171–183 (2005).
- Simson, J. V. & Spicer, S. S. Ferritin particles in macrophages and in associated mast cells. *J. Cell Biol.* **52**, 536–541 (1972).
- Igyártó, B. Z., Lacko, E., Olah, I. & Magyar, A. Characterization of chicken epidermal dendritic cells. *Immunology* **119**, 278–288 (2006).
- Winklhofer, M. & Kirschvink, J. Does avian magnetoreception rely on both magnetite and maghemite? <http://arxiv.org/abs/0805.2249> (2008).
- Walker, M. M. *et al.* Structure and function of the vertebrate magnetic sense. *Nature* **390**, 371–376 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank M. Busslinger, M. Wild and J. Flint for their critical comments on earlier drafts of this manuscript. Thanks also to T. Iancu who commented on our electron micrographs and S. Soto who remarked on the inflammatory lesion in P199. Gratitude is owed to the bio-optics and electron microscopy facilities at the Institute of Molecular Pathology for their assistance in performing experiments. We wish to acknowledge the Centre for Microscopy, Characterisation and Analysis and the Australian Microscopy and Microanalysis Research Facility at the University of Western Australia, a facility funded by the University, State and Commonwealth Governments. Finally, we wish to thank Boehringer Ingelheim, which funds basic science at the Institute of Molecular Pathology.

Author Contributions D.A.K. and C.D.T. conceived and designed the study. M.C.S., C.D.T., M.B., P.P., C.S. and N.E., performed the sectioning, PB staining and counting. D.A.K., C.D.T. and H.C. analysed the resultant data. J.R. and M.L. performed the MRI and CT studies, producing the three-dimensional structure of the pigeon beak. D.A.K. performed the immunohistochemical studies. C.D.T. performed the ultrastructure experiments and J.S. and M.S. did the EFTEM and SAED studies and analysed the data. D.A.K. wrote the paper, and all authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.K. (keys@imp.ac.at).

METHODS

PB staining. We optimized our PB-staining protocol on our large cohort of pigeons ($n = 172$) that we sourced from seven different lofts. We experimented with a variety of different fixatives, section thickness, and blades before adopting the following protocol. Adult homing pigeons ($n = 12$ Nuremberg cohort, $n = 6$ Vienna cohort) were killed before perfusion with 4% phosphate-buffered paraformaldehyde (PFA, pH 7.4). The tissue was dissected with ceramic-coated tools and post-fixed for 18 h before dehydration in an increasing alcohol series. The tissue was then embedded in paraffin and sectioned coronally (10 μm) with ceramic-coated microtome blades (DuraEdgeHigh Profile, BLM00103P). Four consecutive coronal sections were mounted on positively charged microscope slides (Menzel Superfrost PLUS, Thermo Scientific). For PB staining we incubated every third slide in a freshly prepared solution of 5% potassium hexacyanoferrate (Sigma, P9387) in 10% HCl for 20 min. After three washes in double distilled H_2O , sections were counterstained for 2 min in NFR (Sigma, 60700). Each slide was then scanned (MIRAX Slide Scanner) and PB-positive cells counted manually, using light microscopy where necessary. For anatomical mapping, landmarks were identified (Supplementary Fig. 5), and 500- μm -thick increments determined. To obtain estimated total cell counts the number of PB-positive cells counted within a normalized increment were divided by the number of sections counted within that increment and multiplied by a factor of 50 (see Supplementary Tables 1 and 2). We compared the number of PB-positive cells in males and females by performing a Student's *t*-test. All pigeons were sexed using genetic methods as previously described³⁰, and experiments were performed in accordance with the relevant guidelines and regulations (Magistrat 60, Veterinäramt, MA60-001603/2010/002).

Immunohistochemistry. For staining with neuronal antibodies slides were deparaffinized and washed in PBS (pH 7.4) before incubation with the primary antibody for 18 h in 0.1% Triton PBS with 2% milk (pH 7.4). Primary antibodies were used at the following concentrations: NF (Millipore, MAB1621, 1:2,000), TUBB3 (Covance, MMS-435P, 1:1,000), MAP1B (Santa Cruz, SC-58784, 1:75). Following a series of washes in PBS, slides were incubated for 2 h with a biotinylated secondary antibody (1:500), before visualization with a peroxidase-based Vectastain Elite ABC kit (Vector Labs, PK-4002) and the chromophore DAB (3,5-diaminobenzidine, Dako). For MHCII staining, 10- μm cryosections were prepared, quenched in 2% H_2O_2 in PBS for 30 min, before incubation for 18 h in 0.1% Triton PBS with 4% milk (pH 7.4) with the primary antibody (Santa Cruz, SC-59323, 1:500). This antibody is a mouse monoclonal antibody raised against white blood cells originating from the chicken. To avoid cross reaction with endogenous biotin/avidin, a HRP-conjugated secondary antibody was used (Biorad, 1:500), and staining visualized with DAB. Sections were thoroughly washed with PBS before PB staining and scanning as described above. All cell counting and co-localization studies were performed blinded to the antibody used. The overall percentage of co-localization was determined by calculating the rate of co-localization per bird, and ascertaining the mean.

Ultrastructure studies. Adult pigeons were perfused with 2.5% glutaraldehyde supplemented with 2% PFA in PBS (pH 7.4) (Glut-PFA), before tissue dissection with ceramic-coated tools. Following post-fixation for 48 h, this tissue was washed with PBS, incubated for 1 h in phosphate-buffered 2% osmium (pH 7.4), dehydrated and embedded in epoxy resin. After polymerization, the blocks were trimmed and sectioned, alternatively taking semi-thin (2 μm) and ultra-thin (70 nm) sections. The ultra-thin sections were mounted on formvar-film copper slot grids for TEM, whereas the semi-thin sections were etched with 21% sodium ethoxide in ethanol, rehydrated and stained with PB and NFR. Where necessary, PB staining was intensified by incubating the sections in 0.5% DAB³¹. TEM imaging on ultra-thin sections used a 100 kV electron microscope (FEI Morgagni 268D) with a CCD camera (Morada Olympus-SIS). For EFTEM imaging and selected area electron diffraction, the ultra-thin sections were mounted on holey carbon-film copper finder grids (Quantifoil, R3.5/1) and analysed on a 200 kV TEM (JEOL, 2100) fitted with a Gatan Imaging Filter (Tridiem) and CCD camera (Orion SC1000). For EFTEM, bright-field images were taken before obtaining elemental maps for iron, which were acquired using the iron M-edge and generated using the conventional three-window method. Two pre-edge (background) images were acquired at energies of 45 and 50 eV, and the post-edge (signal energy) image was acquired by centring the filter's energy-selecting slit at 59 eV with a slit width of 5 eV (~10 s acquisition time). Diffraction data was obtained from iron deposits identified by EFTEM and the data calibrated against a polycrystalline gold standard.

MRI. Animals were killed and perfused with 4% PFA as described above. Following 18 h of post-fixation, heads were mounted in a 70-mm diameter PE tube filled with proton-free perfluoro-polyether fomblin (Solvay Solexis S.p.A.). Imaging was performed on a horizontal bore 9.4 T DirectDrive VNMRS system (Agilent Technologies) using a 72 mm quadrature birdcage volume coil (RAPID Biomedical GmbH). For three-dimensional imaging, a gradient-echo sequence with the following parameters was used: time to echo (TE), 2.8 ms; time to repetition (TR), 280 ms; flip angle, 40°; six averages, field of view, 70 × 70 × 35 mm; matrix size, 512 × 512 × 256. For higher-resolution images of the beak, PFA-fixed beaks were incubated with 8 mM gadolinium solution (Magnevist, Bayer AG) in PBS for 48 h followed by embedding in agar containing 8 mM gadolinium. Gradient-echo images were acquired using the following parameters: TE, 2.7 ms; TR, 25 ms; flip angle, 45°; five averages; field of view, 28 × 28 × 35 mm; matrix size, 560 × 560 × 700. Regions of interest were segmented using thresholds with manual adjustments where necessary using Amira visualization software (v.5.2.2, Visage Imaging).

CT imaging. CT imaging was performed on a Nucline Nano SPECT/CT imaging system (Mediso) using the following imaging parameters: 360 projection, pitch 0.5, 55 kVp, 145 μA acquired at 45 μm isotropic reconstructed to 50 μm isotropic.

30. Horng, Y. M., Wu, C. P., Wang, Y. C. & Huang, M. C. A novel molecular genetic marker for gender determination of pigeons. *Theriogenology* **65**, 1759–1768 (2006).
31. Moos, T. & Mollgard, K. A sensitive post-DAB enhancement technique for demonstration of iron in the central nervous system. *Histochemistry* **99**, 471–475 (1993).

Unexpected features of *Drosophila* circadian behavioural rhythms under natural conditions

Stefano Vanin^{1*†}, Supriya Bhutani^{2*†}, Stefano Montelli^{1†}, Pamela Menegazzi^{1†}, Edward W. Green², Mirko Pegoraro^{1,2}, Federica Sandrelli¹, Rodolfo Costa¹ & Charalambos P. Kyriacou²

Circadian clocks have evolved to synchronize physiology, metabolism and behaviour to the 24-h geophysical cycles of the Earth¹. *Drosophila melanogaster*'s rhythmic locomotor behaviour provides the main phenotype for the identification of higher eukaryotic clock genes^{2,3}. Under laboratory light–dark cycles, flies show enhanced activity before lights on and off signals^{4,5}, and these anticipatory responses have defined the neuronal sites of the corresponding morning (M) and evening (E) oscillators^{6,7}. However, the natural environment provides much richer cycling environmental stimuli than the laboratory, so we sought to examine fly locomotor rhythms in the wild. Here we show that several key laboratory-based assumptions about circadian behaviour are not supported by natural observations. These include the anticipation of light transitions, the midday 'siesta', the fly's crepuscular activity, its nocturnal behaviour under moonlight, and the dominance of light stimuli over temperature. We also observe a third major locomotor component in addition to M and E, which we term 'A' (afternoon). Furthermore, we show that these natural rhythm phenotypes can be observed in the laboratory by using realistic temperature and light cycle simulations. Our results suggest that a comprehensive re-examination of circadian behaviour and its molecular readouts under simulated natural conditions will provide a more authentic interpretation of the adaptive significance of this important rhythmic phenotype. Such studies should also help to clarify the underlying molecular and neuroanatomical substrates of the clock under natural protocols.

Drosophila melanogaster provides a prominent model system in higher eukaryotes for studying the molecular and neurogenetic basis of circadian behavioural rhythms². The fly's 24-h rhythmic locomotor activity has morning (M) and evening (E) components, interpreted in numerous laboratory studies to reflect the fly's anticipation of regular changes in light–dark transitions^{4,5} (Fig. 1a). Neurogenetic dissection has revealed discrete M and E neuronal clusters that determine these circadian landmarks^{6,7}. We wondered how the complex naturally cycling geophysical environment might affect these behavioural patterns, so we investigated fly locomotor rhythms in wild-type and clock mutant strains under natural conditions, by placing activity monitors outdoors for three seasons (2007–2009), from April to November in Leicester, UK (latitude 52° 38' N), and Treviso, Italy (45° 65' N) (Supplementary Fig. 1).

Generally, and throughout the seasons, clock mutants show very high levels of rhythmicity (85–100%) that reflect the cycling environment driving (entraining) their behaviour, except *Clk^{rk}*, which shows decreased rhythmicity (40%) at lower temperatures (Supplementary Table 1). Figure 1b–d illustrates locomotor profiles of a natural wild-type strain (WTALA) and clock mutants from summer observations in Treviso. Contrary to laboratory-based results, at warmer temperatures the 'siesta' (decreased activity during the hottest parts of the day;

Fig. 1a)^{8–10} was not observed. Instead, and in addition to the well-established M and E locomotor peaks, a prominent mid-afternoon component, which we term 'A', was present in more than 50% of wild-type flies at mean daytime temperatures of 20 °C, rising to about 100% at 27 °C (Fig. 1c). Laboratory studies in light–dark cycles at constant high temperatures of 29 °C (refs 8, 9), or in rectangular light and temperature cycles, do not reveal A components (Fig. 1a). This previously unobserved behaviour was seen in all genotypes at warmer temperatures (Fig. 1d and Supplementary Fig. 2) and may represent a stress/escape response that is phase-locked to the rise and fall in temperature. Alternatively, but not exclusively, A may reflect an environmentally modulated circadian phenotype. We observed that the A component was significantly advanced by up to 3 h in short-period *per^S* mutants in comparison with other *per* genotypes; *per⁰¹* was also advanced compared with *per^L* and the wild type, and the A offset was significantly delayed in *per^L* compared with the other genotypes (Fig. 1e and Supplementary Fig. 2), suggesting that A is modulated by the clock and is not a simple escape reflex.

Under laboratory 12 h light–12 h dark cycles, wild-type flies anticipate lights-on by increasing their locomotor activity about 2 h before the lights-on signal, in contrast with arrhythmic mutants such as *per⁰¹*, which do not anticipate the signal^{6,7}. Consequently, in nature, if wild-type flies anticipate a dawn-related geophysical transition, clock mutants would be expected to 'not anticipate' and delay their activity accordingly. In nature, astronomical, nautical and civil twilights, defined by the angles of the Sun below the horizon (18°, 12° and 6°, reflecting illumination of approximately 0.001, 0.01 and about 1 lx, respectively), accompany the dawn–dusk transitions¹¹. We therefore used nautical twilight, which occurs roughly 1.5–2 h before dawn, as a convenient environmental marker by which to compare the onset of morning activity (M_{onset}) among genotypes. Seasonally, the average daily M_{onset} for WTALA flies in Leicester and Treviso reveals that the two locations differ in their overall profiles (Supplementary Fig. 3a). However, when the values are replotted against average night temperature, this difference evaporates, revealing a similar inverse relationship between M_{onset} and temperature in three wild-type strains (Fig. 2a–c), clock mutants and Gal4-driven genotypes (Fig. 2d–j and Supplementary Fig. 4). The exception is *tim⁰¹*, for which M_{onset} is only marginally temperature-dependent (Fig. 2d) and is rescued by transformation with the *tim⁺* gene, and in *per⁰¹;tim⁰¹* double mutants (Fig. 2e, f). Ablating the Pigment Dispersing Factor (PDF)-expressing (Morning) neurons with *UAS-hidUAS-rpr*, or using the *UAS-CycA* dominant-negative to stop the clock in these cells, failed to disrupt the relationship between M_{onset} and temperature (Supplementary Fig. 4).

We investigated whether clock mutants 'anticipate' changes in twilight differently from the wild type, mirroring laboratory studies^{6,7}. We used an analysis of covariance (ANCOVA; temperature as covariate) to

¹Department of Biology, University of Padova, Padova 35131, Italy. ²Department of Genetics, University of Leicester, LE1 7RH, UK. [†]Present addresses: Department of Chemical and Biological Sciences, School of Applied Science, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK (S.V.); Department of Cellular and Molecular Neurosciences, National Brain Research Centre, Manesar NH8, Haryana, 122050, India (S.B.); Department of Experimental Veterinary Science, University of Padova, Viale dell'Università 16, 35020 Legnaro (Agripolis) Padova, Italy (S.M.); Lehrstuhl für Neurobiologie und Genetik, Universität Würzburg, Biozentrum, Am Hubland, 97074 Würzburg, Germany (P.M.).

*These authors contributed equally to this work.

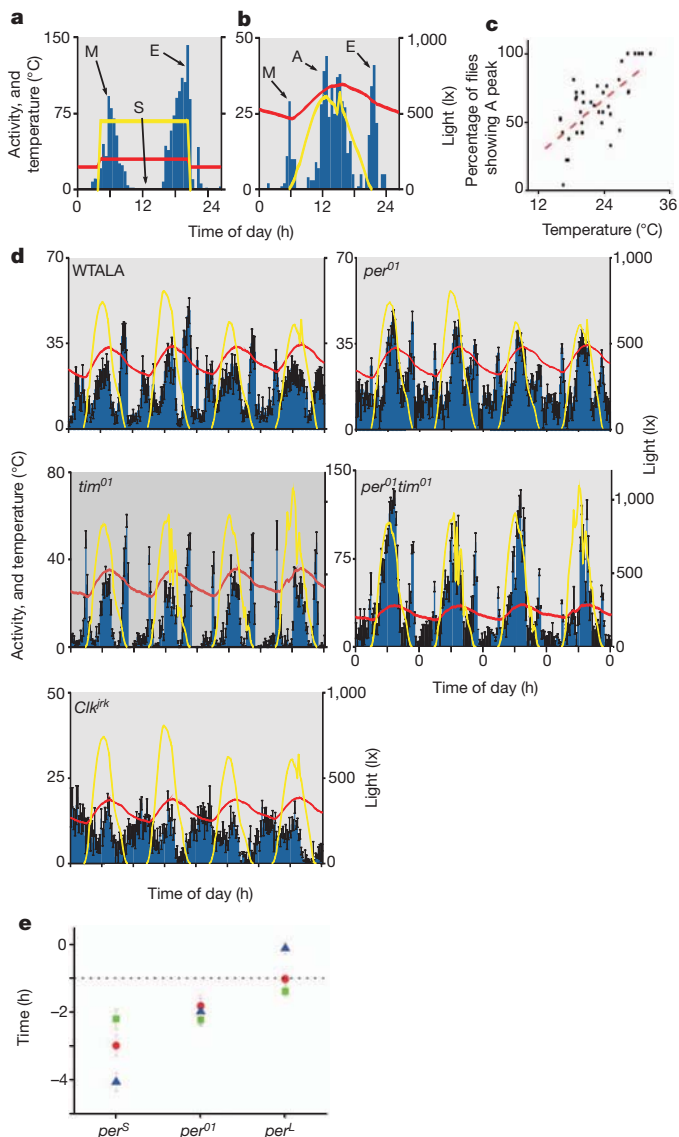


Figure 1 | The afternoon (A) component of locomotor activity. **a**, An individual male's (WTALA) locomotor activity in laboratory conditions with 16 h light–8 h dark and 22 °C in the dark phase and 30 °C in the light. M, morning; S, siesta; E, evening (yellow, light intensity; red, temperature). **b**, An individual male's (WTALA) locomotor activity on a single summer's day (yellow, light intensity; red, temperature; day temperatures: average 28.7 °C, maximum 34.4 °C, minimum 23.4 °C). M, morning; A, afternoon; E, evening. Time 0 = midnight (00:00). **c**, Afternoon peak is associated with temperature ($R^2 = 0.34$, $F_{1,32} = 17.7$; $P < 0.0002$). **d**, Averaged locomotor plots (means \pm s.e.m.) for WTALA and four arrhythmic mutants (see Supplementary Fig. 2 for other mutants). WTALA ($n = 19$), per^{01} ($n = 12$) and Clk^{irk} ($n = 27$, with 3 arrhythmic flies) were monitored together (average temperature over experiment 28.0 °C, maximum 34.4 °C, minimum 21.1 °C), whereas tim^{01} ($n = 17$) and the double mutant $per^{01};tim^{01}$ ($n = 12$) were monitored at a different time (average temperature 29.3 °C, maximum 35.9 °C, minimum 23.1 °C). Red, temperature; yellow, light intensity. **e**, Afternoon component in per mutants compared with wild-type ALA (WTALA). Times of A_{onset} (green squares), A_{peak} (red circles) and A_{offset} (blue triangles) of WTALA were subtracted from the corresponding values for the mutants because not all genotypes were run in the same experiment. Negative values indicate an earlier phenotype than WTALA. ANOVAs for all parameters gave significant genotype effects ($F_{2,151} = 4.8$ (onset), 8.9 (peak) and 29.8 (offset); all $P < 0.001$). In *post hoc* tests, per^S was significantly earlier in all parameters than per^L , and significantly earlier than per^{01} in peak and offset. per^L was significantly delayed compared with per^{01} in all parameters.

inspect the average time from nautical twilight that each clock variant began morning activity, but only Pdf^0 , which was slightly delayed,

fell significantly outside the range of the three wild types (Fig. 2i and Supplementary Table 2a). These results suggest that clock genotypes (except tim^{01}) generally anticipate geophysical variables equally. Alternatively, if M_{onset} activities are simply temperature-sensitive responses to dim light levels, they should be delayed on the following morning if the activity monitors are covered up after the previous sunset. When we performed this experiment, significant delays of about 1 h were observed for all genotypes (except tim^{01}), with M_{onset} now tracking the morning rise in temperature (Fig. 2k, l, Supplementary Fig. 5 and Supplementary Table 2b). The attenuated tim^{01} M_{onset} twilight-modulated temperature responses (Fig. 2d and Supplementary Fig. 5a) recall the mutant's disrupted diapause, a photoperiodic phenotype also requiring both light and temperature input¹². We conclude that M_{onset} is a twilight-dependent temperature response with little evidence for circadian regulation.

We next examined evening onset (E_{onset}), which is seasonally delayed relative to the photoperiod during the warmer summer months (Supplementary Fig. 3b). We replotted E_{onset} against average daytime temperature, relative to the time of maximum daily temperature (T_{max} ; Fig. 3 and Supplementary Fig. 6). All genotypes showed E_{onset} activities that began successively later as mean daily temperatures rose above 20 °C (Fig. 3 and Supplementary Fig. 6). At less than 20 °C, the response was stable in relation to T_{max} , with marginally significant negative relationships for WTALA, HU and Clk^{irk} strains (Supplementary Table 3a) and with E_{onset} generally occurring within the 2.25 ± 0.17 h (mean \pm s.e.m.) that defined the time between the maximum daily light intensity and the subsequent T_{max} (Fig. 3a–i). When analysed separately for temperatures below and above 20 °C, per^S showed the earliest E_{onset} activities, followed by per^{01} and tim^{01} ; per^L had the latest (Supplementary Table 3b, c). This is reflected in Fig. 3, where the data points fall either mostly above (later, per^L) or below (earlier, per^S , per^{01} , tim^{01}) the upper dotted line representing T_{max} . These results mirror laboratory findings in which per^S and per^L modulate the timing of evening behaviour¹³; in addition, however, the unexpected observation of E components in arrhythmic mutants, and their phase advance, indicates an underlying residual short-period rhythmicity^{5,14–16}.

In laboratory wild-type flies, the length of the temperature-sensitive afternoon 'siesta' is mediated by the thermal regulation of per 3' splicing^{8–10}. E_{onset} defines the end of the siesta, so we studied whether the per splicing readout correlated with the E_{onset} curve. Levels of per 3' splicing in fly heads throughout the seasons for WTALA flies (Fig. 3j) were linearly related to temperature over the entire range 7–30 °C, so it is unlikely that splicing contributed significantly to E_{onset} at temperatures below 20 °C (Supplementary Fig. 6F). Indeed, because the laboratory siesta is mediated by the dynamics of the upswing of the clock protein PERIOD (PER)⁸, the fact that per^{01} -null mutants showed a similar E_{onset} –temperature relationship to that of the wild type also argues against a significant role for per 3' splicing in this natural phenotype. A similar conclusion using splice-locked per transformants was reached in laboratory experiments using artificial temperature and light cycles¹⁷.

Drosophila rhythms in the laboratory are often described as 'crepuscular' because of the distribution of M and E components at dawn and dusk¹⁸. We therefore quantified the activity falling in the morning and evening twilights (astronomical, nautical and civil) as a proportion of all activity falling between the beginning of morning and the end of evening astronomical twilights. Using data from Italian experiments, we observed that even at the warmest temperatures when M_{onset} and E_{onset} are pushed towards the twilights (Figs 1–3), the proportion of activity falling within the twilights was less than 25% and was reduced still further at cooler temperatures (Fig. 3k). Thus, in nature the major proportion of activity, even at high temperatures, falls outside the twilights, so flies are diurnal rather than crepuscular.

Related laboratory experiments have also suggested that locomotor activity becomes predominantly crepuscular and nocturnal under

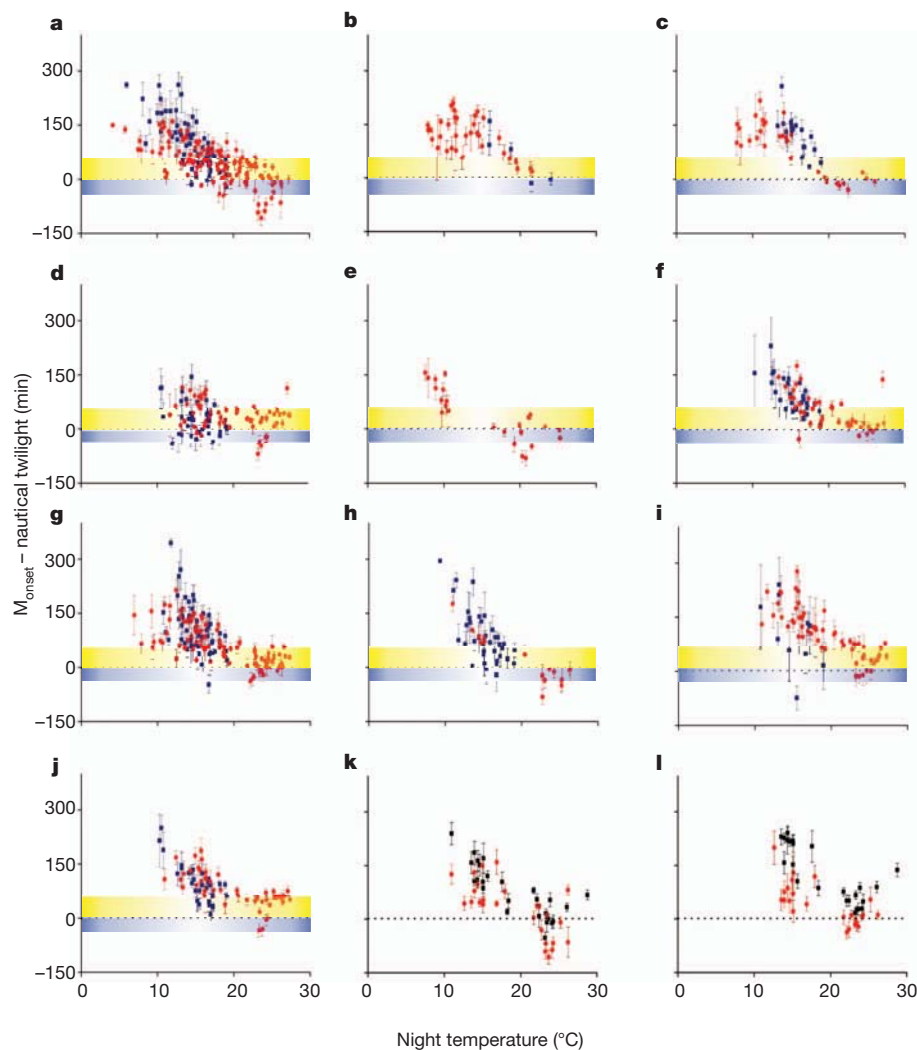


Figure 2 | The morning onset (M_{onset}) of locomotor activity is dependent on temperature and twilight. **a–j**, Time of M_{onset} in a single day compared with nautical twilight (dotted line, time 0). Blue and yellow areas represent the approximate time between astronomical and nautical twilight and between nautical twilight and sunrise, respectively (both calculated as seasonal averages for Treviso). Each point represents the mean of 4–24 flies (median 14); error bars represent s.e.m. Blue squares, Leicester; red circles, Treviso. **a–c**, Wild

types: WTALA (Alto Adige) (**a**), Canton S (**b**) and HU (Houten) (**c**). **d–j**, Clock mutants: tim^{01} (**d**), $tim^{01};P[tim^{+}]$ (**e**), $per^{01};tim^{01}$ (**f**), per^{01} (**g**), Clk^{jrk} (**h**), Pdf^0 (**i**) and cry^{01} (**j**) (Supplementary Fig. 4). The slope for tim^{01} is significantly different from that for all three wild types (Supplementary Table 2a). **k, l**, M_{onset} of wild type (**k**) and per^{01} mutant (**l**) before (red) and after (black) activity monitors were covered at sunset and maintained until mid-afternoon (Supplementary Fig. 5 and Supplementary Table 2b).

simulated moonlight and this correlates with delayed expression of the clock proteins PER and TIMELESS (TIM) in the fifth PDF-negative sLNv neuron^{19,20}. We calculated the proportion of WTALA nocturnal activity during new moons and immediately before and after full moons. Figure 4a reveals no significant differences in nocturnal activity between the two conditions, even when we varied the exposure to moonlight from 0.03 to 0.33 lx (Supplementary Fig. 7). When we analysed PER in adult brains by immunocytochemistry under conditions of full moon and no moon, consistent with the behaviour, PER expression in the fifth sLNv (Fig. 4b) was not significantly delayed in moonlight in comparison with the other lateral clock neurons (s-LNvs, l-LNvs and LNDs). The dorsal DN1 and DN2 neurons showed significantly earlier PER peaks than the other neurons (Fig. 4b and Supplementary Fig. 8), indicative of a faster underlying oscillation²¹. These results suggest that natural temperature and light cycles may counteract the delayed PER expression of the fifth sLNv observed under simulated moonlight.

Finally, we attempted to reproduce natural circadian phenotypes within the laboratory by simulating cycling temperature and light intensity. We simulated consecutive days of new moon (darkness at night), with astronomical twilight (0.005 lx) giving way gradually to

nautical twilight (0.05 lx), to civil twilight (1 lx) over a 2-h period at the beginning and (in reverse) at the end of the day, rising to a maximum intensity of 1,500 lx in the afternoon. We ran this at cycling day–night temperatures ranging from 20 to 30 °C (spring) or from 25 to 35 °C (summer), mimicking actual temperatures that we measured in two Treviso spring and summer days, respectively, when the photoperiod for both was roughly 16 h light–8 h dark. We observed that in the summer simulation, M, A and E components were prominently displayed by all genotypes (WTALA, per mutants, tim^{01} and Clk^{jrk}), although Clk^{jrk} showed a smaller E burst on most days (Fig. 4c and Supplementary Fig. 9). We confirmed that M_{onset} occurs between the simulated times of astronomical and nautical twilight, as in the wild at these temperatures, with all genotypes becoming active 0.2–0.7 h before simulated summer nautical twilight, but later (between 0.1 and 0.4 h) before simulated spring twilight (Supplementary Fig. 9c). As in the wild, M_{onset} is temperature-sensitive, with no significant differences between the wild type and either per^{01} or tim^{01} mutants in the timing on M_{onset} (although per^S and per^L were slightly earlier in their M_{onset} for both spring and summer simulations), and with tim^{01} showing a similar temperature independence to that in the wild (Supplementary Fig. 9c). We observed the circadian modulation of

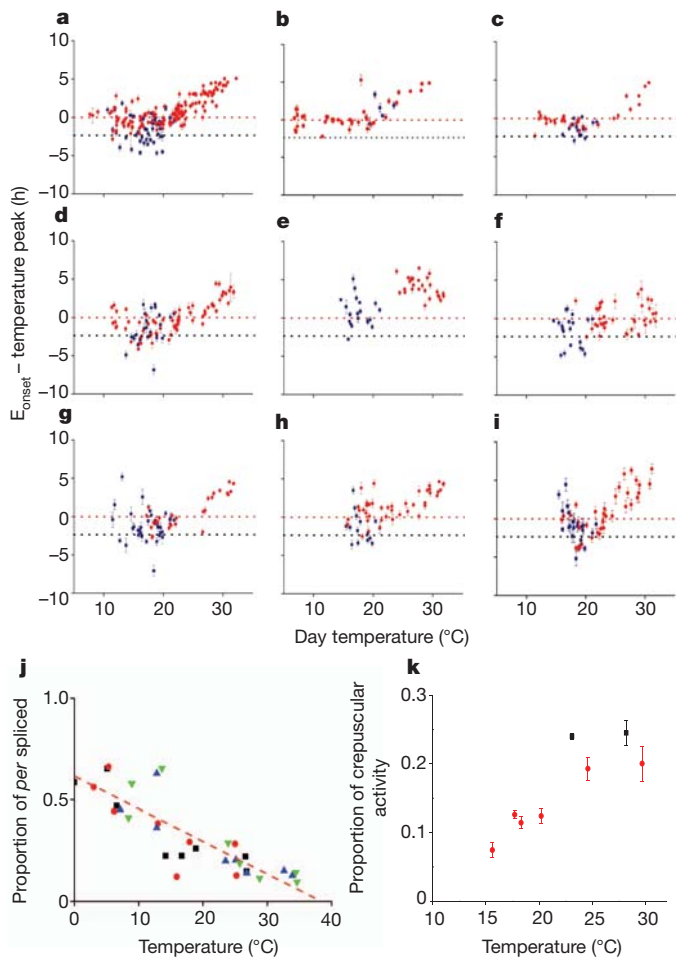


Figure 3 | Evening onset is temperature and clock modulated. **a–i**, Evening onset compared with daily maximum temperature (T_{\max} , 0). Dotted line below T_{\max} represents the average time that light peaked (about 2.25 h) before T_{\max} . Blue squares, Leicester; red circles, Treviso. **a–c**, Wild types: WTALA (Alto Adige) (a), Canton S (b) and HU (Houten) (c). **d–i**, Clock mutants: *per⁰¹* (d), *per^L* (e), *per^S* (f), *tim⁰¹* (g), *per⁰¹;tim⁰¹* (h) and *Clk^{rk}* (i) (see also Supplementary Fig. 6 and Supplementary Table 3). Most data points below 20 °C fall between the times of the light and temperature peaks. *per⁰¹*, *tim⁰¹* and *per^S* have earlier E_{onset} (ANCOVA: $P \approx 0$, 0.003 and 0.0007, respectively, versus WTALA) as shown by the lower scores on the y axis; *per^L* has later onset ($P = 0.0003$ versus wild-type). **j**, *per* splicing in nature in fly heads. *per^{spliced}* isoform compared with *per^{total}* (*per^{spliced}* + *per^{unspliced}*) against average daily temperature in WTALA strain. Heads collected at 00:00 (black squares), 03:00 (red circles), 12:00 (blue triangles) and 15:00 (green triangles). **k**, Locomotor activity is neither crepuscular in natural conditions nor in simulated natural conditions. Red circles, mean \pm s.e.m. of daily crepuscular activity in natural condition of groups of WTALA flies in Italy from single multi-day experiments; black squares, Italian spring and summer laboratory simulations (see Fig. 4c).

the E component in our simulations (Supplementary Fig. 9e) and whereas nearly all flies, irrespective of clock genotype, showed a prominent A component in the summer simulation, this proportion fell to 20–50% in the spring simulation, matching our results from nature and revealing the importance of temperature for the expression of A (Fig. 4c and Supplementary Fig. 9b, d). We also observed similar proportions of crepuscular activity for the spring and summer simulations to those in our natural observations (Fig. 3k). We did not, however, reproduce the advance in the A component of *per^S* compared with the wild type that we observed in nature (Supplementary Fig. 9d).

Our results from the wild reveal a diminished role for ‘anticipation’ of light–dark transitions partly because arrhythmic mutants also show M and E components, and also because the M component appears to be a response to changes in the environment. However, the newly

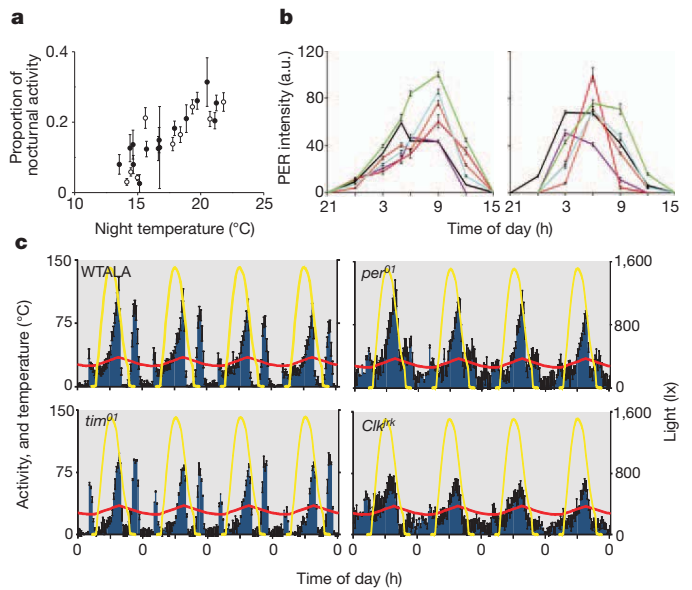


Figure 4 | Diurnal phenotypes of *D. melanogaster*. **a**, The proportion of nocturnal locomotor activity of WTALA males does not differ between days of full moon (open circles) and no moon (filled circles) (ANCOVA: $F_{1,22} = 0.04$, $P = 0.84$). Results are means \pm s.e.m. and are based on 4–24 flies (median 17). **b**, PER immunofluorescence intensity (mean \pm s.e.m., arbitrary units) in clock neurons in May 2008 (left) and September 2008 (right) (new and full moon, respectively, corresponding to filled and open circles in a). Red, fifth Pdf-null sLNv; brown, sLNvs; green, LNDs; turquoise, l-LNvs. DN1 neurons (black) and DN2 neurons (purple) are significantly advanced ($F_{5,498} = 13.71$, $P \approx 0$; see also Supplementary Fig. 8). There is a significant advance for all neurons in September in comparison with May ($F_{5,498} = 34.1$, $P \approx 0$) but no month \times neuron interaction ($F_{5,498} = 0.79$, $P = 0.56$). May, lights on at 05:16, lights off at 20:59, photoperiod 15 h 43 min; September, lights on at 06:20, lights off at 19:50, photoperiod 13 h 30 min. **c**, Simulation of Italian summer in the laboratory for the wild type and for arrhythmic mutants (16 h light–8 h dark, temperature range 25–35 °C; see also Supplementary Fig. 9). $n = 19$ –32 for each genotype at each temperature.

described A and the E components are clock-modulated, unlike M. This may be because M occurs during the rapid changes in illumination from quasi-darkness to bright light, when temperature is constant at its daily minimum, whereas A and E components occur when the light intensity is peaking or falling rapidly but temperature is either still rising or falling gradually. Such conflicting environmental signals may have recruited the clock to regulate appropriately adaptive A and E locomotor responses. Our results also show that temperature is the critical variable for predicting circadian behaviour¹⁷, because the large differences in photoperiod between our two locations (more than 3 h in midsummer) contribute little to the behavioural variance. In contrast, laboratory studies that place temperature and light cycles in different phases to each other²² reveal that light is the most important driver of circadian behaviour.

Although lateral and dorsal circadian neurons underlying M and E behaviour in light–dark cycles at constant temperature have been identified^{6,7}, subsets of dorsal neurons and lateral posterior neurons may be relevant for circadian temperature signalling^{22–26}. Perhaps in the wild it is these neurons that dominate the neuronal circuit. However, our observation that various ‘clockless’ genotypes can nevertheless show M and E (and A) components also suggests that the underlying neuronal substrates for this behavioural programme may not reside solely in the lateral and dorsal neurons but also elsewhere in the brain. Clock gene expression enhances and modulates circadian behaviour in the wild, but the possibility remains that a residual short-period rhythmicity resides within *per* and *tim* ‘arrhythmic’ mutants^{5,14–16} that can be amplified under natural conditions. Our results resonate with studies of laboratory mice²⁷ and hamsters²⁸,

which show them to behave quite differently in nature, where they are predominantly diurnal and not nocturnal as in the laboratory. Similarly, generally accepted adaptive and mechanistic explanations for fly circadian behaviour from laboratory experiments may require some revision if they are to account for rhythmicity in nature.

METHODS SUMMARY

Fly strains. Strains WTALA and HU are natural isolates collected as isofemale lines from northern Italy and The Netherlands¹². Canton-S is a standard laboratory strain. *per*⁰¹, *per*^S and *per*^L are congenic and Cantonized, but the genetic backgrounds of the other mutants used *per*⁰¹; *tim*⁰¹, *Clk*^{rk}, *Pdf*⁰, *cry*⁰ and Gal4-driven transgenes (drivers *tim-gal4* and *Pdf-gal4*), the apoptotic *UAS-hidUAS-rpr* and the dominant-negative *UAS-CycA* (ref. 29) are not equivalent, and for most are unknown.

Behaviour. Flies were raised at 25 °C in 12 h light–12 h dark cycles in the laboratory before males were placed outside at 2–3 days of age in Trikinetics activity monitors. Morning (*M*_{onset}) and afternoon or evening (*A*_{onset}, *E*_{onset}) onsets were determined operationally for each fly on each day of every experiment. Circadian rhythmicity was assessed by spectral analysis and autocorrelation as reported previously³⁰.

Moonlight measurements. Measurements of moonlight were made in Treviso and Leicester, using a LI-210SA Photometric Sensor (LI-COR) connected to a LI 1400 data logger.

Twilights. Twilights were obtained from the online database of the United States Naval Observatory (USNO) Astronomy Application Department¹¹.

Immunocytochemistry. The slides were examined under a Nikon 80i microscope with QICAM Fast Camera. Individual images were taken of planes at different depths to create a z-series for each brain. Brains were stained with primary antibodies against PDF and PER (the latter was a gift from R. Stanewsky).

RNA quantification. Flies were collected every 3 h for a 24-h time course in natural conditions, and fixed in liquid nitrogen. Heads were removed and *per*^{spliced}, *per*^{unspliced} forms and the control gene *Cbp20* (*Cap binding protein 20*) were amplified by PCR. PCR products were revealed on a 2% agarose gel and imaged, and final quantification was obtained with ImageJ software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 July 2011; accepted 24 February 2012.

Published online 4 April 2012.

- Dodd, A. N. *et al.* Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**, 630–633 (2005).
- Hardin, P. E. Molecular genetic analysis of timekeeping in *Drosophila*. *Adv. Genet.* **74**, 141–173 (2011).
- Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nature Rev. Mol. Cell Biol.* **8**, 139–148 (2007).
- Helfrich-Forster, C. The locomotor activity rhythm of *Drosophila melanogaster* is controlled by a dual oscillator system. *J. Insect Physiol.* **47**, 877–887 (2001).
- Collins, B. H., Dissel, S., Gaten, E., Rosato, E. & Kyriacou, C. P. Disruption of Cryptochrome partially restores circadian rhythmicity to the arrhythmic period mutant of *Drosophila*. *Proc. Natl Acad. Sci. USA* **102**, 19021–19026 (2005).
- Grima, B., Chelot, E., Xia, R. & Rouyer, F. Morning and evening peaks of activity rely on different clock neurons of the *Drosophila* brain. *Nature* **431**, 869–873 (2004).
- Stoleru, D., Peng, Y., Agosto, J. & Rosbash, M. Coupled oscillators control morning and evening locomotor behaviour of *Drosophila*. *Nature* **431**, 862–868 (2004).
- Majercak, J., Sidote, D., Hardin, P. E. & Edery, I. How a circadian clock adapts to seasonal decreases in temperature and day length. *Neuron* **24**, 219–230 (1999).
- Collins, B. H., Rosato, E. & Kyriacou, C. P. Seasonal behavior in *Drosophila melanogaster* requires the photoreceptors, the circadian clock, and phospholipase C. *Proc. Natl Acad. Sci. USA* **101**, 1945–1950 (2004).
- Low, K. H., Lim, C., Ko, H. W. & Edery, I. Natural variation in the splice site strength of a clock gene and species-specific thermal adaptation. *Neuron* **60**, 1054–1067 (2008).
- United States Naval Observatory (USNO) Astronomy Application Department. *Rise, Set, and Twilight Definitions* (http://aa.usno.navy.mil/faq/docs/RST_defs.php) (2011).
- Tauber, E. *et al.* Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* **316**, 1895–1898 (2007).
- Hamblen-Coyle, M. J., Wheeler, D. A., Rutla, J. E., Rosbash, M. & Hall, J. C. Behavior of period-altered rhythm mutants of *Drosophila* in light:dark cycles. *J. Insect Behav.* **5**, 417–446 (1992).
- Helfrich, C. & Engelmann, W. Evidences for circadian rhythmicity in the *per*⁰ mutant of *Drosophila melanogaster*. *Z. Naturforsch. C* **42**, 1335–1338 (1987).
- Dowse, H. B. & Ringo, J. M. Further evidence that the circadian clock in *Drosophila* is a population of coupled ultradian oscillators. *J. Biol. Rhythms* **2**, 65–76 (1987).
- Yoshii, T., Fujii, K. & Tomioka, K. Induction of *Drosophila* behavioral and molecular circadian rhythms by temperature steps in constant light. *J. Biol. Rhythms* **22**, 103–114 (2007).
- Currie, J., Goda, T. & Wijnen, H. Selective entrainment of the *Drosophila* circadian clock to daily gradients in environmental temperature. *BMC Biol.* **7**, 49 (2009).
- Rieger, D. *et al.* The fruit fly *Drosophila melanogaster* favors dim light and times its activity peaks to early dawn and late dusk. *J. Biol. Rhythms* **22**, 387–399 (2007).
- Bachleitner, W., Kempinger, L., Wulbeck, C., Rieger, D. & Helfrich-Forster, C. Moonlight shifts the endogenous clock of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **104**, 3538–3543 (2007).
- Kempinger, L., Dittmann, R., Rieger, D. & Helfrich-Forster, C. The nocturnal activity of fruit flies exposed to artificial moonlight is partly caused by direct light effects on the activity level that bypass the endogenous clock. *Chronobiol. Int.* **26**, 151–166 (2009).
- Netabach, M. N. *et al.* Electrical hyperexcitation of lateral ventral pacemaker neurons desynchronises downstream circadian oscillators in the fly circadian circuit and induces multiple behavioural period. *J. Neurosci.* **26**, 479–489 (2006).
- Miyasako, Y., Umezaki, Y. & Tomioka, K. Separate sets of cerebral clock neurons are responsible for light and temperature entrainment of *Drosophila* circadian locomotor rhythms. *J. Biol. Rhythms* **22**, 115–126 (2007).
- Yoshii, T., Vanin, S., Costa, R. & Helfrich-Forster, C. Synergic entrainment of *Drosophila*'s circadian clock by light and temperature. *J. Biol. Rhythms* **24**, 452–464 (2009).
- Zhang, Y., Liu, Y., Bilodeau-Wentworth, D., Hardin, P. E. & Emery, P. Light and temperature control the contribution of specific DN1 neurons to *Drosophila* circadian behavior. *Curr. Biol.* **20**, 600–605 (2010).
- Zhang, L. *et al.* DN1(p) circadian neurons coordinate acute light and PDF inputs to produce robust daily behavior in *Drosophila*. *Curr. Biol.* **20**, 591–599 (2010).
- Busza, A., Murad, A. & Emery, P. Interactions between circadian neurons control temperature synchronization of *Drosophila* behavior. *J. Neurosci.* **27**, 10722–10733 (2007).
- Daan, S. *et al.* Lab mice in the field: unorthodox daily activity and effects of a dysfunctional circadian clock allele. *J. Biol. Rhythms* **26**, 118–129 (2011).
- Gatterman, R. *et al.* Golden hamsters are nocturnal in captivity but diurnal in nature. *Biol. Lett.* **4**, 253–255 (2008).
- Tanoue, S., Krishnan, P., Krishnan, B., Dryer, S. E. & Hardin, P. E. Circadian clocks in antennal neurons are necessary and sufficient for olfaction rhythms in *Drosophila*. *Curr. Biol.* **14**, 638–649 (2004).
- Rosato, E. & Kyriacou, C. P. Analysis of locomotor activity rhythms in *Drosophila*. *Nature Protocols* **1**, 559–568 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Bastianello for building the natural light simulator, and C. Helfrich-Forster for comments on the manuscript. This work was funded by grants from the European Community (the 6th Framework Project EUCLOCK no. 018741; to R.C. and C.P.K.), the Biotechnology and Biological Sciences Research Council and National Environmental Research Council (to C.P.K.), the Royal Society Wolfson Research Merit Award (to C.P.K.) and a Medical Research Council studentship (to E.W.G.). R.C. also thanks the Italian Space Agency (DCMC grant) and the Ministero dell'Università e delle Ricerche (MIUR).

Author Contributions C.P.K. and R.C. conceived the study. C.P.K. supervised the UK work, and R.C. the Italian component. S.B. conducted all the field experiments in the UK, and S.V., F.S. and M.P. the fieldwork in Italy. S.V. performed the laboratory simulations. S.M. performed the *per* splicing work, and P.M. the immunocytochemistry experiments. E.W.G. developed the software to analyse the behavioural data. Statistical analyses were performed by S.B., S.V. and C.P.K. C.P.K., S.V., S.B. and R.C. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.C. (rodolfo.costa@unipd.it).

METHODS

Fly strains. Strains WTALA and HU are natural isolates collected as isofemale lines from northern Italy (Alto Adige 46° 30' N) in 2004 and from Houton, The Netherlands (52° 2' N) in 2004, respectively. Canton-S is a standard laboratory strain. *per*⁰¹, *per*^S and *per*^L are congenic and were backcrossed repeatedly to a Cantonized *per* deficiency *Df(1)64j4* background by C.P.K. in the early 1980s. The genetic backgrounds of the other fly strains used are varied, and for most they are unknown.

In the laboratory, even at a constant 10 °C, we observed that *UAS-CycA* (ref. 29) was effectively driven by the *timgal4* driver and gave an arrhythmic locomotor profile compared to the control strain (Supplementary Fig. 10). All flies were raised to adulthood at 25 °C in the laboratory before being placed outside at 2–3 days of age. We confirmed that *UAS-hidUAS-rpr* was effective in eliminating the clock neurons under *tim*- and *Pdf-gal4* control using immunocytochemistry (data not shown).

Locomotor activity. Locomotor activity was monitored in suburban gardens well away from street lighting and from any major roads. Flies were placed in Trikinetics activity monitors, and shielded from direct sunlight and from rain (Supplementary Fig. 1). In Leicester this was accomplished by placing the monitors on a table within a child's plastic playhouse (Toys R Us) with approximate internal dimensions of 4 × 4 × 4 ft³, with open windows in a north–south direction that was shaded by an adjacent large tree. Bricks placed on the roof ensured that the house did not blow away during high winds. Cables from monitors were fed from the back of the playhouse into a computer that was placed inside an immediately adjacent extension to C.P.K.'s house. In Italy, monitors were placed under a penthouse roof in a garden roughly 4 km from the city centre of Treviso, again with no direct sunlight, with cables fed to a computer within the penthouse. Flies were monitored from March to November (Treviso) or from May to September (Leicester). Simultaneous and continuous monitoring of temperature, light intensity and humidity were made with recorders placed adjacent to the activity tubes. Circadian rhythmicity was assessed by spectral analysis and autocorrelation as reported previously³⁰.

Morning (*M*_{onset}) and afternoon or evening (*A*_{onset}, *E*_{onset}) onsets were determined for each fly on each day of every experiment (see Supplementary Fig. 11 for examples). *M*_{onset} was considered to be present when a bout of activity occurred after a period of rest during the dark phase, if continuous movement with a steady increase in activity (and no more than one half-hour time bin without any or lower activity than the previous bin) led to a peak, followed by a steady decrease in activity defining the offset. Activity bouts in the middle of the night (not leading continuously into the morning) were not considered. In general, the window of time that contained the morning bout was 3 h before or after the 30-min time bin during which the light intensity reached 1 lx. If there was no bout of activity consistent with the parameters described above, it was noted that there was no morning component for that fly on that day.

*E*_{onset} was considered to be present when a bout of activity occurred after a period of rest during the day. If a burst of activity occurred in the middle of the afternoon on warmer days, with another clear bout after it, the latter was considered to represent the *E*_{onset}, whereas the former represented the afternoon component. The evening and afternoon bouts of activity had to be composed of continuous movement with no more than one zero activity bin interspersed within, and with a steady increase and decrease in activity levels defining the onset/offset. If there was no bout of activity consistent with the parameters described above, it was noted that there was no evening peak for that fly on that day.

Moonlight measurements. Measurements of moonlight were made in Treviso and Leicester, using a LI-210SA Photometric Sensor (LI-COR) connected to a LI 1400 data logger. Five-minute readings were made four times during selected nights of full moon or new moon, and the minimum and maximum light levels were recorded, both in fully exposed full moonlight conditions and in the more sheltered positions in which the Trikinetics monitors were normally maintained. Average maximum nightly full moonlight levels were 0.23 ± 0.15 lx in fully exposed positions, in comparison with 0.07 ± 0.01 lx in sheltered conditions in Leicester (30 November 2009). The corresponding values in Treviso were 0.33 ± 0.07 lx in exposed positions and 0.03 ± 0.01 lx in sheltered conditions. At new moon, in both Leicester and Treviso, the light values were 0 lx.

Twilights. Twilights were obtained from the online database of the United States Naval Observatory (USNO) Astronomy Application Department³¹.

Natural temperature profiles. These were reproduced in the laboratory by using the Memmert ICP 700 incubator with a refrigeration unit. The Celsius 2007 computer program, running under the Windows XP operating system, was used for programming, controlling and documenting the Memmert ICP 700 using the USB interface. The incubator can generate smooth cycling temperature ranges without step-ups.

Natural light profiles. These were simulated by using a programmable daylight simulator based on a battery of white-light LEDs of different intensities that could be programmed to generate smooth cycling profiles (without step-ups) of light ranging from 0.005 lx (astronomical twilight) to 1,500 lx. The front-end programming allowed the simulation of any observed natural cycle of light intensity. The equipment was designed and built by EURITMI, a spin-out from the Electronic Engineering Unit at the Venetian Institute of Molecular Medicine (Padova, Italy).

Immunocytochemistry. Flies were collected every 3 h for a 24-h time course in natural conditions and fixed for 3 h in 4% paraformaldehyde. After three washes (15 min each) in PBS, the brains were dissected under a stereomicroscope (Leica MB6). Brains were washed five times for 6 min each in 0.3% PBST (PBS + 0.3% Triton X-100). They were then permeabilized for 10 min in 1% PBST and blocked for 2 h with 1% BSA. After blocking, the brains were incubated for 3 days at 4 °C, in the primary antibody diluted in 0.1% BSA in 0.3% PBST. They were then washed five times for 6 min each in 0.3% PBST. After washing, they were blocked again for 1 h with 1% BSA. The brains were then incubated overnight at 4 °C in the appropriate secondary antibody. The antibodies used for the immunocytochemistry experiments were anti-PDF (Developmental Studies Hybridoma Bank, dilution of 1:5,000), anti-PER (from R. Stanewsky; 1:2,500), and anti-mouse and anti-rabbit (both from Invitrogen; 1:500).

Brains were mounted onto slides (VWR) with a drop of mounting medium (Vectashield; Vecta Laboratories, Inc.) and covered with a coverslip 0.1 mm thick (VWR). The slides were observed under a Nikon 80i microscope with a QICAM Fast Camera. Individual images were taken of planes at different depths to create a *z*-series for each brain. The size of the sections forming a *z*-series was 1.0 ± 0.2 µm. The images were viewed and quantified with ImageJ version 1.42g (<http://rsb.info.nih.gov/ij/>). The average pixel intensity for each neuron was measured together with the signal from its corresponding background area. The final amount of signal was calculated using the formula intensity = 100 × (signal – background)/background. The values of intensity obtained were normalized using the higher value as 100% and then plotted with Microsoft Excel 2003 or OriginPro 8.0, and statistical analyses were performed with Statistica 8 (StatSoft).

RNA quantification. Flies were collected every 3 h for a 24-h time course in natural conditions and fixed in liquid nitrogen. Heads were removed and total RNA was recovered and extracted using Trizol Reagent (Gibco) as recommended by the kit protocol. Reverse transcription was initiated on total RNA, after any contaminating genomic DNA was removed with DNase (Promega), using the SuperscriptII (Invitrogen) reverse transcriptase with the 17-bases oligo(dT). The reaction was performed for 1 h at 42 °C and for 15 min at 75 °C. *per*^{spliced}, *per*^{unspliced} forms and the control gene *Cbp20* (*Cap binding protein 20*) were amplified by PCR with a PTC-100 Peltier Thermal Cycler (MJ Research) using primers listed below. PCR was performed as follows: initial denaturation at 95 °C for 3 min, then 28 cycles consisting of 95 °C for 1 min, 62.1 °C for 1 min, and 72 °C for 45 s. The reaction was completed by an elongation step of 10 min at 72 °C. Amplifications were carried out in 20-µl reaction mixtures containing 75 ng of cDNA target, 1 µl of each primer (10 µM), 1.6 µl of dNTPs (2 mM), 4 µl of Green Buffer (5×) and 0.4 µl of GoTaq DNA polymerase (Promega). Primers used were as follows: *per*, 5'-AAGACGAGCCGGCTCCAG-3' (forward; base pairs 6421–6440; NCBI X03636) and 5'-TCTACATTATCCTCGGCTTGC-3' (reverse; base pairs 7201–7221; NCBI X03636); *cbp20*, 5'-GTCTGATTCGTGTGGACTGG-3' (forward; base pairs 540–559; FlyBase ID FBgn0022943) and 5'-CAACAGTTTGCCATAACCC-3' (reverse; base pairs 653–672; FlyBase ID FBgn0022943).

PCR products were revealed on a 2% agarose (Eurobio) gel under ultraviolet radiation. Images were collected with the Quantity One 4.6 (Bio-Rad). The final quantification was obtained with ImageJ software by applying the formula intensity_{RNA(pixel)} = band intensity of *per* spliced or *per* unspliced/band intensity of *cbp20*. Each band's intensity was corrected for background noise. For each collection, three independent RNA extractions, reverse transcriptions, amplifications and quantifications were performed. The results were normalized and analysed with Statistica 8 (StatSoft). The plots were obtained with OriginPro 8.0 software.

Local generation of glia is a major astrocyte source in postnatal cortex

Woo-Ping Ge¹, Atsushi Miyawaki², Fred H. Gage³, Yuh Nung Jan¹ & Lily Yeh Jan¹

Glial cells constitute nearly 50% of the cells in the human brain¹. Astrocytes, which make up the largest glial population, are crucial to the regulation of synaptic connectivity during postnatal development². Because defects in astrocyte generation are associated with severe neurological disorders such as brain tumours³, it is important to understand how astrocytes are produced. Astrocytes reportedly arise from two sources^{4–6}: radial glia in the ventricular zone and progenitors in the subventricular zone, with the contribution from each region shifting with time. During the first three weeks of postnatal development, the glial cell population, which contains predominantly astrocytes, expands 6–8-fold in the rodent brain⁷. Little is known about the mechanisms underlying this expansion. Here we show that a major source of glia in the postnatal cortex in mice is the local proliferation of differentiated astrocytes. Unlike glial progenitors in the subventricular zone, differentiated astrocytes undergo symmetric division, and their progeny integrate functionally into the existing glial network as mature astrocytes that form endfeet with blood vessels, couple electrically to neighbouring astrocytes, and take up glutamate after neuronal activity.

Most radial glia have finished producing their share of astrocytes and have begun to disappear shortly after birth^{4–6}; astrocytes are therefore thought to derive mainly from progenitors in the subventricular zone (SVZ) at later stages⁸. The massive expansion of glia within the first three postnatal weeks presents a daunting task for their presumed SVZ progenitors. This task is rendered even more challenging by the thickening of the cortex compounded by the disappearance of radial glia, which provides the migratory tracks for newly formed astrocytes⁹. We used electroporation to transfect green fluorescent protein (GFP) plasmids into SVZ/radial glial cells of mice at postnatal days (P)0–2 to label them with GFP *in vivo* and to trace their progeny at P16–20 (Fig. 1a and Supplementary Fig. 1). Only a very small percentage (about 3%) of the astrocytes derived postnatally from SVZ/radial glial cells reached cortical layers I–IV; most were left behind in SVZ/white matter (75%) and layers V–VI (22%) (Fig. 1b, c). It therefore seems that huge numbers of cortical astrocytes generated postnatally might arise from a more efficient process, such as local cell proliferation (Ki67⁺; Fig. 1d), rather than from SVZ progenitors. Whereas glial cell division within the cortex was reported half a century ago, on the basis of studies involving [³H]thymidine incorporation into DNA¹⁰, the extent of the contribution of local glial division to postnatal astrocyte production remained unknown, owing to the difficulty in distinguishing glia generated locally from glia derived from other sources. In this study, we have obtained evidence to support the hypothesis that the local generation of astrocytes within the postnatal cortex is a major source of glia.

To label locally generated glia, we used a replication-defective murine leukaemia retrovirus (MLV) to express GFP in infected dividing cells and their progeny in postnatal cortex *in vivo*. This retrovirus specifically infects proliferating cells and has been used for cell-fate tracing in SVZ and the hippocampal subgranular zone (SGZ) *in vivo*^{8,11}. We injected

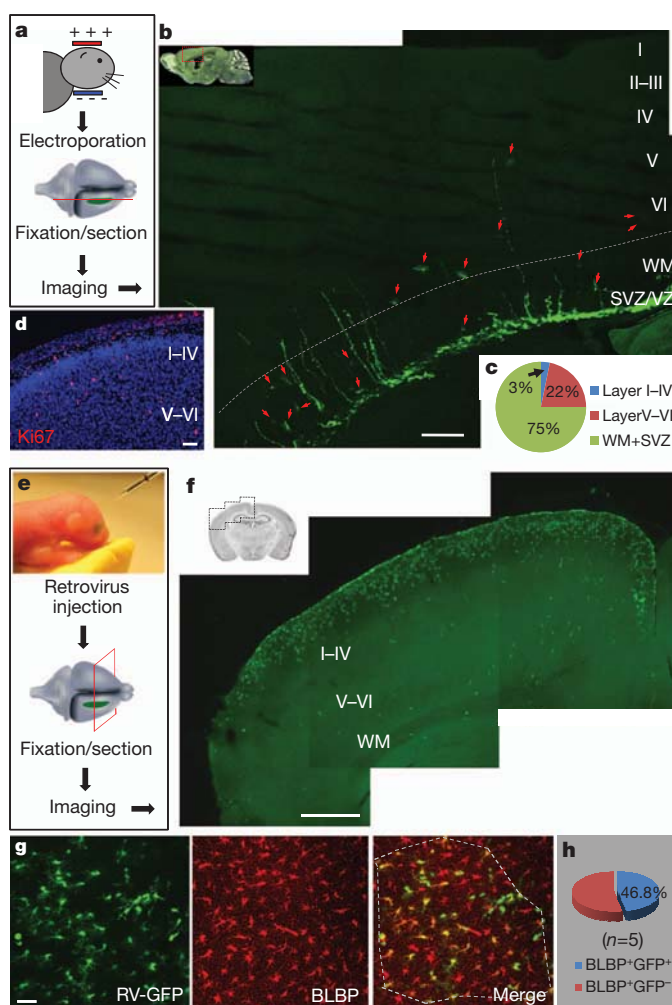


Figure 1 | Locally generated glia as a major source of astrocytes.

a, Procedure to label SVZ/radial glia-derived astrocytes by electroporation. **b**, The distribution of astrocytes (arrows) 2 weeks after electroporation. VZ, ventricular zone. **c**, Percentages of astrocytes at different locations. WM, white matter. **d**, Proliferating cells (Ki67⁺, red) in a cortical section of P3 mouse. Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, blue). **e**, Procedure to label locally proliferating cells by retrovirus. **f**, Cells labelled by retrovirus (green). **g**, Image of infected astrocytes. Astrocytes (BLBP⁺, red) with GFP (GFP⁺BLBP⁺) or without GFP (GFP⁻BLBP⁺) in the outlined region (dashed line) were included for analysis in **h**. RV, retrovirus. **h**, Percentages of astrocytes labelled by retrovirus injected locally, calculated as 100 × (BLBP⁺GFP⁺ cells/BLBP⁺ cells). Scale bars, 200 μm (**b**), 50 μm (**d**), 500 μm (**f**) and 40 μm (**g**).

¹Howard Hughes Medical Institute, Departments of Physiology, Biochemistry and Biophysics, University of California at San Francisco, 1550 4th Street, San Francisco, California 94158, USA. ²Life Function and Dynamics, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, and Brain Science Institute, RIKEN, Wako-city, Saitama, 351-0198, Japan. ³Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

viruses locally into layers I–IV of the barrel or motor cortex of wild-type mice at P0–6 (Fig. 1e, f) and examined GFP expression 1 week later in samples that were also stained with antibodies against brain lipid-binding protein (BLBP) (Fig. 1g), which labels radial glia during embryonic development and astrocytes in the postnatal brain¹². Whereas about 30% of infected cells were NG2 glia (27.6%, $n = 662$ infected cells, Supplementary Fig. 2), 55–70% of infected cells were astrocytes (BLBP⁺, 56.9%, $n = 369$ GFP⁺ cells, Fig. 1g; GFAP⁺, 68.6%, $n = 662$ GFP⁺ cells, Supplementary Fig. 2), indicating that these astrocytes originated locally *in vivo*.

To determine whether a major astrocyte source was derived from the local generation of glia, we injected retroviruses with higher titre ($1 \mu\text{l}$, $(1-3) \times 10^7$) into the cortex of P0–2 mice and compared the number of GFP-expressing astrocytes (BLBP⁺GFP⁺) with the total number of astrocytes (BLBP⁺) within an infected region 7–10 days

after injection. We found that nearly half of the astrocytes were doubly labelled (GFP⁺BLBP⁺, $46.8 \pm 3.8\%$, $n = 5$ mice; Fig. 1f–h). Because the half-life of infectivity of MLV retrovirus is 5–8 h at 37 °C, the doubly labelled astrocytes probably correspond to astrocytes undergoing division in the time window of 5–8 h plus the progeny they generate over the course of 7–10 days. Control studies revealed no differences in the morphology of astrocytes (Fig. 1g), the density of dividing cells (Supplementary Fig. 3) or the percentage of GFAP-occupied area (Supplementary Fig. 4) in brain regions with or without retroviral infection. Our observations therefore suggest that local proliferation is a major source of astrocytes in the postnatal cortex.

To test the possibility that multiple dividing cell types infected by retroviruses, including astrocytes, NG2 glia and perhaps some unknown progenitors in the cortex, gave rise to these GFP-expressing astrocytes, we labelled acute brain slices with the nuclear marker

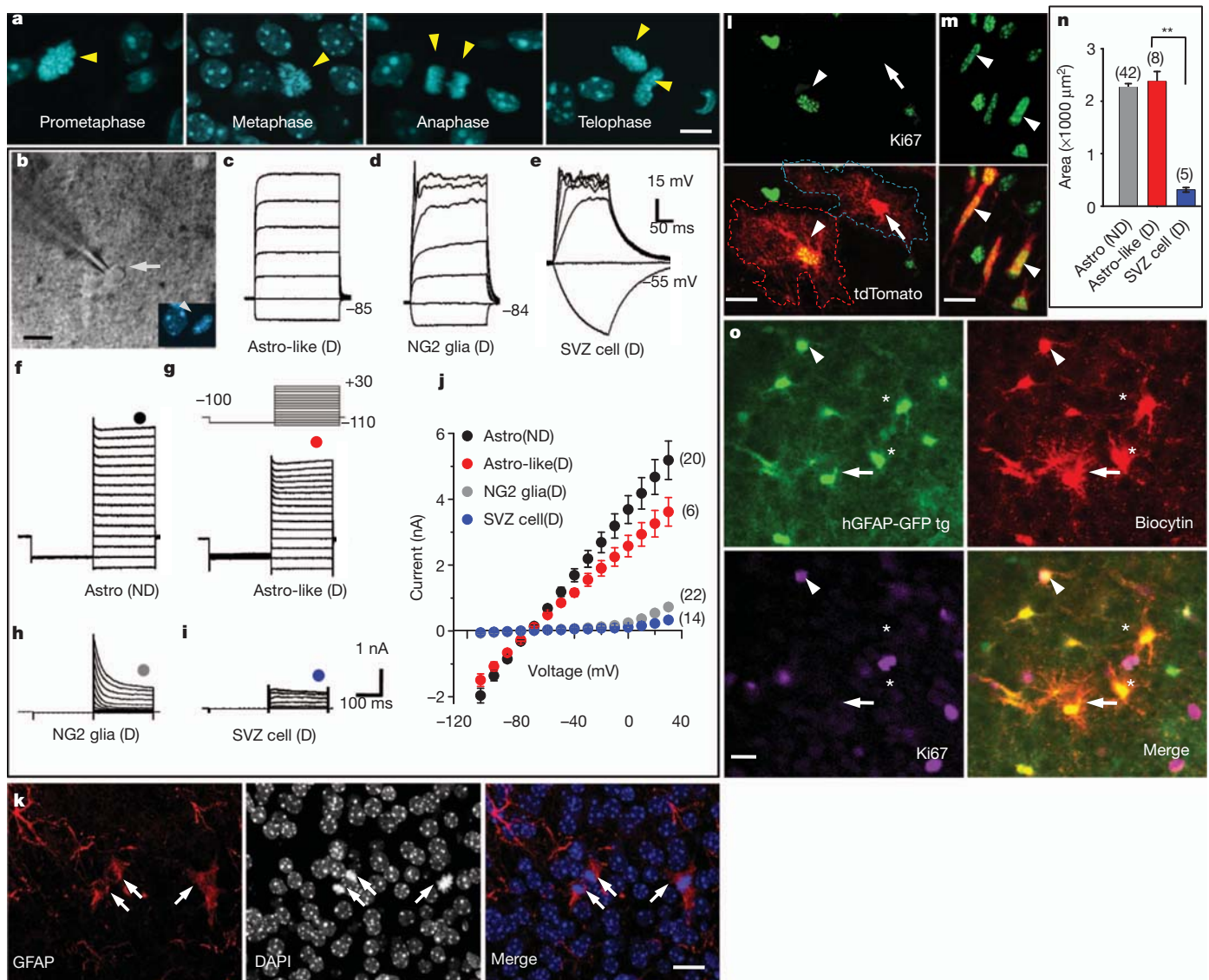


Figure 2 | Properties of dividing cells within the cortex. **a**, Nuclei (arrowheads, Hoechst 33342) of dividing cells at different mitotic stages in acute brain slices. **b**, A dividing cell (differential interference contrast, arrow) at prometaphase (Hoechst 33342, arrowhead). **c–e**, Voltage responses from an Astro-like-D cell (**c**), a dividing NG2 glia (**d**) and a dividing SVZ cell (**e**). **f–i**, Current responses from a non-dividing (ND) astrocyte (**f**), an Astro-like-D cell (**g**), a dividing NG2 glia (**h**) and a dividing SVZ cell (**i**) evoked by step voltages (inset, **g**). **j**, Current–voltage curves in **f–i** (the circles indicate the positions of measurements). **k**, Astro-like-D cells (arrows; two at telophase, one at metaphase) stained with anti-GFAP (red). **l, m**, Morphology of a non-dividing astrocyte (arrow) and an Astro-like-D cell (arrowhead, **l**) in the cortex, and dividing cells (arrowheads) in the SVZ (**m**) of a P8 hGFAP-CreER;Ai14 transgenic mouse. **n**, Summary of the area covered by the processes of non-dividing astrocytes (grey), Astro-like-D cells (red) and SVZ dividing cells (blue) (10- μm z-projection with soma included, one-way analysis of variance followed by a Bonferroni post-hoc test; two asterisks, $P < 0.01$). **o**, A non-dividing astrocyte (arrow) was loaded with biocytin. A dividing astrocyte is labelled by biocytin (arrowhead). tg, transgenic. Scale bars, 5 μm (**a**) and 10 μm (**b, k–m, o**). Error bars indicate s.e.m.

dividing astrocyte (arrow) and an Astro-like-D cell (arrowhead, **l**) in the cortex, and dividing cells (arrowheads) in the SVZ (**m**) of a P8 hGFAP-CreER;Ai14 transgenic mouse. **n**, Summary of the area covered by the processes of non-dividing astrocytes (grey), Astro-like-D cells (red) and SVZ dividing cells (blue) (10- μm z-projection with soma included, one-way analysis of variance followed by a Bonferroni post-hoc test; two asterisks, $P < 0.01$). **o**, A non-dividing astrocyte (arrow) was loaded with biocytin. A dividing astrocyte is labelled by biocytin (arrowhead). tg, transgenic. Scale bars, 5 μm (**a**) and 10 μm (**b, k–m, o**). Error bars indicate s.e.m.

Hoechst 33342, a dye that can permeate live cell membranes, to distinguish dividing cells from non-dividing cells (Fig. 2a). In some experiments we also used slices from CAG-Fucci-Green transgenic mice to identify dividing cells in the SVZ and cortex (Supplementary Fig. 5). In this line, the green fluorescent protein mAG accumulates specifically during the S (synthesis) to M (mitosis) stages of the cell cycle¹³, thus facilitating the identification of dividing cells for whole-cell patch-clamp recordings (Fig. 2b). Excluding cells of the vascular system, 94.6% (87 of 92 cells, P6–13) of the dividing cells in the cortex fell into two groups: dividing NG2 glia¹⁴, with characteristic small sodium currents and rectifying current–voltage (I – V) curve (Fig. 2d, h, j) and astrocyte-like dividing cells (Astro-like-D; Fig. 2c, g, j), so named for their similarity to differentiated astrocytes (Fig. 2f, j). Astrocytes characteristically displayed large, delayed rectifier potassium currents (K_{dr}) and large, inwardly rectifying potassium currents (K_{ir}) but no sodium currents, and they had a linear I – V curve¹⁵ (Fig. 2f, j). In contrast, dividing cells recorded in the SVZ (Supplementary Fig. 5) had no K_{ir} current and a very small K_{dr} (Fig. 2e, i, j), typical of immature progenitors¹⁶. To further characterize the Astro-like-D cells, our immunostaining revealed that they were GFAP⁺ but Nestin[−] (Fig. 2k and Supplementary Fig. 6). We then compared the morphology of

Astro-like-D cells with that of mature astrocytes or SVZ dividing progenitors in hGFAP-CreER;Ai14 transgenic mice. Crossing Ai14 transgenic mice¹⁷ with hGFAP-CreER transgenic mice¹⁸ allowed robust expression of the red fluorescent protein tdTomato after inducible astrocyte-specific Cre-mediated recombination. We administered tamoxifen to hGFAP-CreER;Ai14 transgenic mice at P0–2 and assessed the cellular morphology 1 week later (Supplementary Fig. 7). Astro-like-D cells (Ki67⁺tdTomato⁺) had a complex morphology comparable to that of neighbouring mature astrocytes (Ki67[−]tdTomato⁺; Fig. 2l, n and Supplementary Fig. 8). In contrast, SVZ dividing progenitors had a bipolar/unipolar morphology (Ki67⁺tdTomato⁺; Fig. 2m, n and Supplementary Fig. 8). Because coupling by means of gap junctions is a hallmark feature of astrocytes, we injected biocytin into individual non-dividing astrocytes in brain slices of hGFAP-GFP transgenic mice, which express GFP under the control of human astrocyte-specific GFAP promoter¹⁹, and found that Astro-like-D cells (Ki67⁺GFP⁺) and mature astrocytes (Ki67[−]GFP⁺; Fig. 2o) were coupled by means of gap junctions. Thus, unlike SVZ dividing progenitors (Fig. 2e, i, j) and glioblasts²⁰, Astro-like-D cells in the cortex are differentiated astrocytes. Taken together with a previous report²¹ and our tracing results from NG2-CreBac/ERTM;Ai14

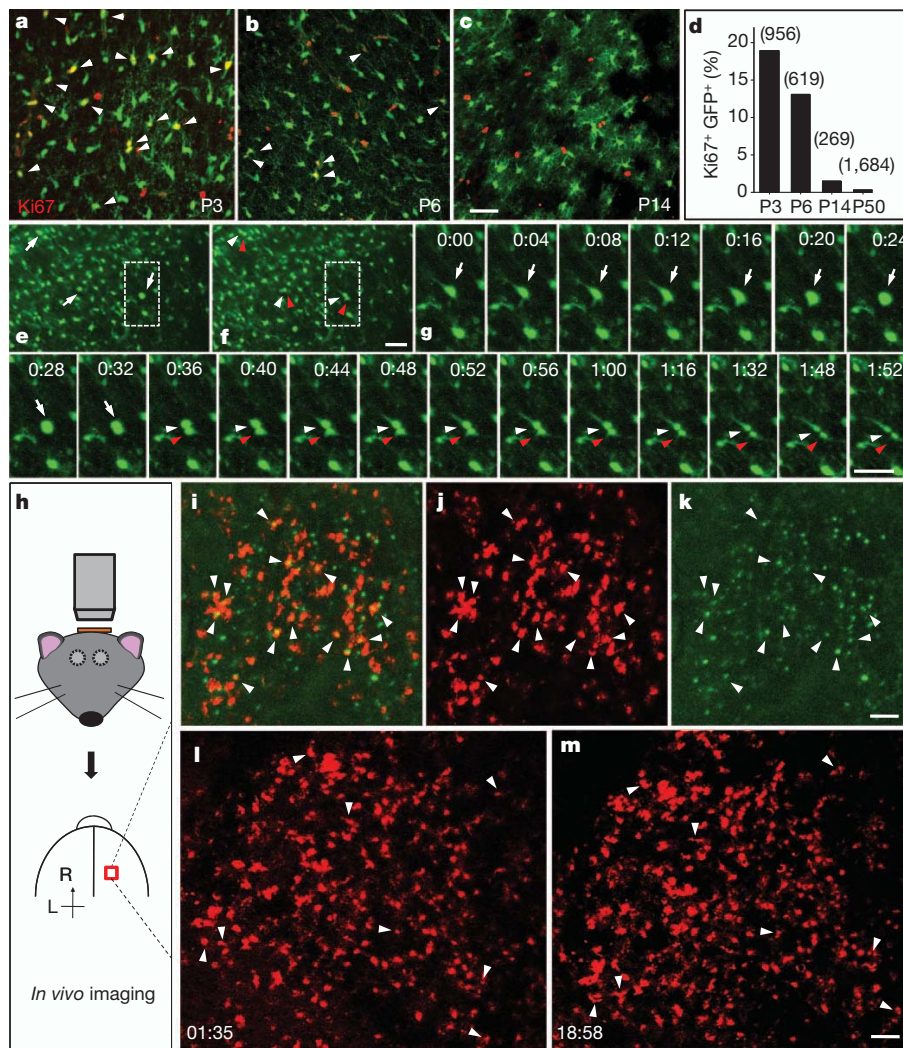


Figure 3 | Time-lapse imaging of local proliferation of astrocytes. a–c, Proliferating astrocytes (arrowheads) in the cortex of hGFAP-GFP transgenic mice, at P3 (a), P6 (b) and P14 (c). d, Summarized data for the percentage of Ki67⁺GFP⁺ cells with strong GFP signals. e, f, Sequential images of a cortical slice from a P3 hGFAP-GFP transgenic mouse (e, parent cells (arrows); f, daughter cells (arrowheads)). g, Time-lapse images

(1 h 52 min) of a dividing GFP⁺ cell in e and f. h, Procedure to image cell division *in vivo*. i–k, Images from a P4 triply transgenic hGFAP-CreER;Ai14;CAG-Fucci-Green mouse (i, combined images; j, tdTomato; k, mAG signal; arrowheads, dividing astrocytes). l, m, Time-lapse images at 1 h 35 min (l) and 18 h 58 min (m) from a P5 hGFAP-CreER;Ai14 transgenic mouse (arrowheads, dividing astrocytes). Scale bars, 40 μ m (a–g) and 100 μ m (k, m).

transgenic mice showing that NG2 glia generated very few astrocytes in the cortex during postnatal life (data not shown), our results reveal that Astro-like-D cells are the parent cells of locally generated astrocytes.

To assess the abundance of proliferating astrocytes within the cortex, we perfused hGFAP-GFP transgenic mice for Ki67 immunostaining and observed numerous astrocytes in the process of cell division before P10 (18.9% at P3, $n = 956$ GFP⁺ cells; 13.1% at P6, $n = 619$ GFP⁺ cells; 1.5% at P14, $n = 269$ GFP⁺ cells; 0.30% at P48–52, $n = 1,684$ GFP⁺ cells; Fig. 3a–d and Supplementary Fig. 9). To directly monitor local generation of astrocytes, we performed time-lapse imaging of acute cortical slices from hGFAP-GFP transgenic mice and found that roughly 2% of astrocytes divided within 3 h ($2.0 \pm 0.2\%$, 16 of 809 cells with strong GFP signals from five mice, P3–5; Fig. 3e–g and Supplementary Movie 1). Because NG2 glia maintain proliferative ability throughout life¹⁴, and some hippocampal NG2 glia are reported to have a very weak GFP signal in another hGFAP-GFP transgenic line²², we tested whether NG2 glia could have been among the dividing cells with a strong GFP signal, by loading biocytin into cells with a strong GFP signal through a recording pipette. We found that biocytin diffused only among cells with strong GFP signals ($n = 5$ slices; Supplementary Fig. 10), which is consistent with our observations from doubly transgenic hGFAP-GFP;NG2BacDsRed mice (Supplementary Fig. 11). Thus, the dividing cells with strong GFP expression are astrocytes rather than NG2 glia.

The density of dividing cells in acute brain slices showed no obvious change within a 6-h period after slice preparation (Supplementary Fig. 12); however, mature astrocytes could conceivably be induced to undergo local gliogenesis by means of a stab wound *in vivo*²³. We therefore performed *in vivo* imaging with an open skull but an intact pial surface within 1 h after surgery on the triply transgenic hGFAP-CreER;Ai14;CAG-Fucci-Green mice (Fig. 3h). We observed abundant dividing astrocytes (12.4%, tdTomato⁺mAG⁺; Fig. 3i–k, P3–6), a similar observation to that in brain sections (Fig. 3a, b, d). Because the thinned skull preparation does not cause astrocytic gliosis or activation of microglia²⁴, we then performed long-term time-lapse imaging with the thinned skull preparation in hGFAP-CreER;Ai14 transgenic mice and obtained similar results: abundant astrocytes were generated locally within the cortex (about 8% of progeny, $(8 \times 2)/212$ cells in Fig. 3l, m). Because of the difficulty of identifying a dividing cell under thinned skull if its two daughter cells did not separate completely, we probably underestimated the percentage of astrocytes produced on the basis of *in vivo* time-lapse imaging.

To determine whether dividing astrocytes in the cortex undergo symmetric division to produce astrocytes, or asymmetric division to generate multiple cell types as SVZ cells do^{8,25}, we recorded from their progeny during and shortly after cytokinesis (Fig. 4a, b). The two daughter cells shared similar *I*–*V* relationships that were characteristic of astrocytes (Fig. 4c). In addition, we examined the daughter-cell morphology in P6–8 hGFAP-CreER;Ai14 transgenic mice and found the two daughter cells occupying comparable areas and showing similar labelling with the astrocyte marker, BLBP (Fig. 4d and Supplementary Table 1). Thus, locally dividing astrocytes in the cortex primarily undergo symmetrical division to generate two daughter astrocytes.

To determine whether the progeny maintained their astrocytic fate after exiting from the cell cycle, we administered tamoxifen at P0–2 to hGFAP-CreER;Ai14 transgenic mice to label astrocytes permanently with tdTomato and examined their locally generated progeny 1 week later ($n = 4$ mice) to test whether these tdTomato⁺ cells still expressed the astrocyte marker BLBP. Although there were many progenitors or neurons with tdTomato expression in SVZ and hippocampal dentate gyrus (Supplementary Fig. 7f, g), nearly all of the tdTomato⁺ cells were BLBP⁺ or GFAP⁺ in the cortex (99.8%, motor and barrel cortex; Supplementary Fig. 7a–e). Because we found that few astrocytes would enter programmed cell death in the cortex (Supplementary Fig. 13), it is most likely that the progeny arising from local astrocyte division retained astrocytic identity long after exiting from the cell cycle. For

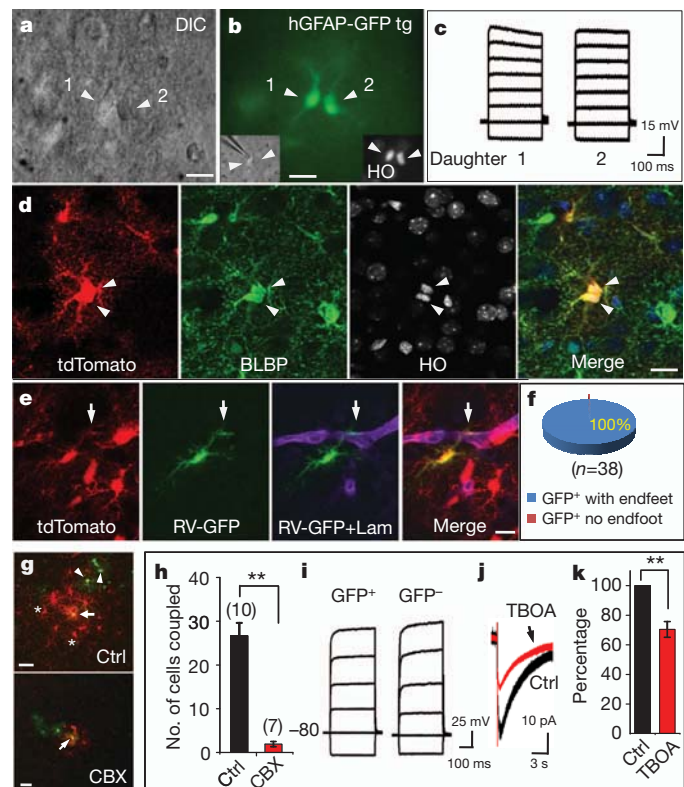


Figure 4 | Symmetric division of proliferating astrocytes and the function of their progeny. **a**, A pair of daughter astrocytes (arrowheads) at late telophase under differential interference contrast. **b**, Both cells had GFP signal. Nuclei were stained with Hoechst 33342 (HO, inset). **c**, Voltage responses of two daughter cells evoked by step currents (–1 to 6 nA). **d**, Two daughter cells in telophase (arrowheads) from a P8 hGFAP-CreER;Ai14 transgenic mouse. **e**, An astrocyte infected by GFP-expressing retroviruses (green) and expressing tdTomato (red) formed endfeet (arrows) with blood vessels (Laminin⁺, purple) in a P19 hGFAP-CreER;Ai14 transgenic mouse. Tamoxifen was injected at P2, and cells were infected with retrovirus at P5. **f**, The percentage of progeny cells marked by retroviruses (GFP⁺tdTomato⁺) that had endfeet (GFP⁺tdTomato⁺ with endfeet, blue). **g**, A retrovirus-infected astrocyte progeny (GFP⁺, green, arrow) in the absence (Ctrl, upper) or presence (lower) of 100 μ M carbenoxolone (CBX) was injected with biocytin (red). Without CBX, both GFP⁺ astrocytes (arrowheads) and GFP[–] astrocytes (asterisks) contained biocytin (red), as a result of gap-junction coupling with the astrocyte progeny injected with biocytin. **h**, The number of cells coupled. Two asterisks, $P < 0.01$, (unpaired *t*-test). **i**, Current responses of uninfected (GFP[–]) and infected (GFP⁺) astrocyte progeny. **j**, **k**, Glutamate transporter current (**j**) and its summarized data (**k**) from infected astrocyte progeny before (black) and after (red) application of blocker TBOA (100 μ M, $70.4 \pm 5.3\%$, $n = 7$). Two asterisks, $P < 0.01$ (paired *t*-test). Scale bars, 10 μ m (**a**, **b**, **d**, **e**) and 20 μ m (**g**). Error bars indicate s.e.m.

further confirmation, tamoxifen was administered at P0–2 and retroviruses were injected locally at P3–5 (3 days after tamoxifen) in the cortex of hGFAP-CreER;Ai14 transgenic mice. The fate of doubly labelled cells (tdTomato⁺GFP⁺) was then assessed 2 weeks later. Because tdTomato marked cells that had expressed astrocyte markers before retroviral infection (Supplementary Fig. 7), the tdTomato⁺GFP⁺ cells correspond to the progeny of astrocytes that were infected by retroviruses during cell division. We found that all doubly labelled cells formed endfoot-like structures with blood vessels (38 of 38 yellow cells; Fig. 4e, f), a characteristic of differentiated astrocytes. These results demonstrate that locally generated progeny retain astrocytic identity long after they exit from the cell cycle.

We then asked whether daughter astrocytes arising from local astrocyte division integrate functionally into the existing glial network as mature astrocytes. The intercellular communication by means of gap junctions between astrocytes has a critical function in ion buffering

in the brain²⁶. At 1–3 weeks after viral infection, we loaded biocytin into infected astrocytes and found that biocytin diffused into neighbouring GFP⁺ astrocytes as well as some GFP⁺ astrocytes within 20 min (26.7 ± 2.9 cells, $n = 10$ slices; Fig. 4g–i and Supplementary Fig. 14). The coupling was inhibited by carbonoxelene (CBX), a blocker for gap junctions (1.9 ± 0.6 cells, $n = 7$ slices; Fig. 4g, h and Supplementary Fig. 14), indicating that locally generated astrocytes successfully integrated into existing glial networks. Another classical function of astrocytes is to clear glutamate from the synaptic cleft²⁷. After stimulation of neuronal fibres, an inward current with slow decay time course appeared in all the GFP⁺ astrocytes recorded (20 of 20 GFP⁺ cells, P12–19; Fig. 4j) and was sensitive to the glutamate transporter blocker TBOA (Fig. 4j, k). The remaining currents, lasting for more than 10 s (decay time 13.3 ± 0.4 s, $n = 4$; Fig. 4j), correspond to K_{ir} activation after neuronal excitation²⁸. These observations thus reveal that locally generated astrocytes function as mature astrocytes to take up glutamate and K⁺ ions after neuronal activity.

We have demonstrated that local generation of astrocytes within the postnatal cortex provides a major glial source, at least in layers I–IV, whereas astrocytes generated early in development are derived from radial glia^{4–6}, SVZ progenitors (including SVZ glioblasts and migratory glioblasts)^{4–6,20}. Once a subset of early astrocytes from those sources colonize and differentiate in the cortex as ‘pioneers’, local division of these differentiated astrocytes has a pivotal role in glial production after birth in the cortex.

Astrocytic endfeet almost fully cover the blood vessels by postnatal day 20 and are crucial to the regulation of cerebral blood flow²⁹ and the transport of nutrients from blood to neurons³⁰. It is not yet clear how this large number of locally generated astrocytes can coordinate with angiogenesis to form the complete gliovascular interface. Furthermore, aberrant gene activity affecting glial proliferation is one potential cause of gliomas, which comprise nearly 80% of primary malignant brain tumours³. It will also be of interest to test whether gliomas could have arisen from defective regulation of locally dividing glial cells in the brain.

METHODS SUMMARY

For live nuclear labelling, slices were incubated with Hoechst 33342 ($2 \mu\text{g ml}^{-1}$) as described previously¹⁴. For *in vivo* imaging, the pups with thinned skull (Fig. 3h, l, m) were immobilized with 4% agarose. The mouth of the pup was attached to a 1-ml pipette tip that was connected to a tube for inhalation. All data are given as means \pm s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 August 2011; accepted 14 February 2012.

Published online 28 March 2012.

1. Azevedo, F. A. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513**, 532–541 (2009).
2. Eroglu, C. & Barres, B. A. Regulation of synaptic connectivity by glia. *Nature* **468**, 223–231 (2010).
3. Schwartzbaum, J. A., Fisher, J. L., Aldape, K. D. & Wrensch, M. Epidemiology and molecular pathology of glioma. *Nat. Clin. Pract. Neurol.* **2**, 494–503 (2006).
4. Cameron, R. S. & Rakic, P. Glial cell lineage in the cerebral cortex: a review and synthesis. *Glia* **4**, 124–137 (1991).
5. Marshall, C. A., Suzuki, S. O. & Goldman, J. E. Gliogenic and neurogenic progenitors of the subventricular zone: who are they, where did they come from, and where are they going? *Glia* **43**, 52–61 (2003).
6. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).
7. Bandeira, F., Lent, R. & Herculano-Houzel, S. Changing numbers of neuronal and non-neuronal cells underlie postnatal brain growth in the rat. *Proc. Natl Acad. Sci. USA* **106**, 14108–14113 (2009).
8. Levison, S. W. & Goldman, J. E. Both oligodendrocytes and astrocytes develop from progenitors in the subventricular zone of postnatal rat forebrain. *Neuron* **10**, 201–212 (1993).
9. Zerlin, M., Levison, S. W. & Goldman, J. E. Early patterns of migration, morphogenesis, and intermediate filament expression of subventricular zone cells in the postnatal rat forebrain. *J. Neurosci.* **15**, 7238–7249 (1995).

10. Smart, I. & Leblond, C. P. Evidence for division and transformations of neuroglia cells in the mouse brain, as derived from radioautography after injection of thymidine-H³. *J. Comp. Neurol.* **116**, 349–367 (1961).
11. Suh, H. *et al.* *In vivo* fate analysis reveals the multipotent and self-renewal capacities of Sox2⁺ neural stem cells in the adult hippocampus. *Cell Stem Cell* **1**, 515–528 (2007).
12. Guo, F., Ma, J., McCauley, E., Bannerman, P. & Pleasure, D. Early postnatal proteolipid promoter-expressing progenitors produce multilineage cells *in vivo*. *J. Neurosci.* **29**, 7256–7270 (2009).
13. Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
14. Ge, W. P., Zhou, W., Luo, Q., Jan, L. Y. & Jan, Y. N. Dividing glial cells maintain differentiated properties including complex morphology and functional synapses. *Proc. Natl Acad. Sci. USA* **106**, 328–333 (2009).
15. Ge, W. P. *et al.* Long-term potentiation of neuron–glia synapses mediated by Ca²⁺-permeable AMPA receptors. *Science* **312**, 1533–1537 (2006).
16. Wang, D. D., Krueger, D. D. & Bordey, A. Biophysical properties and ionic signature of neuronal progenitors of the postnatal subventricular zone *in situ*. *J. Neurophysiol.* **90**, 2291–2302 (2003).
17. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
18. Casper, K. B., Jones, K. & McCarthy, K. D. Characterization of astrocyte-specific conditional knockouts. *Genesis* **45**, 292–299 (2007).
19. Zhuo, L. *et al.* Live astrocytes visualized by green fluorescent protein in transgenic mice. *Dev. Biol.* **187**, 36–42 (1997).
20. Burns, K. A., Murphy, B., Danzer, S. C. & Kuan, C. Y. Developmental and post-injury cortical gliogenesis: a genetic fate-mapping study with Nestin-CreER mice. *Glia* **57**, 1115–1129 (2009).
21. Zhu, X. *et al.* Age-dependent fate and lineage restriction of single NG2 cells. *Development* **138**, 745–753 (2011).
22. Matthias, K. *et al.* Segregated expression of AMPA-type glutamate receptors and glutamate transporters defines distinct astrocyte populations in the mouse hippocampus. *J. Neurosci.* **23**, 1750–1758 (2003).
23. Buffo, A. *et al.* Origin and progeny of reactive gliosis: a source of multipotent cells in the injured brain. *Proc. Natl Acad. Sci. USA* **105**, 3581–3586 (2008).
24. Xu, H. T., Pan, F., Yang, G. & Gan, W. B. Choice of cranial window type for *in vivo* imaging affects dendritic spine turnover in the cortex. *Nature Neurosci.* **10**, 549–551 (2007).
25. Doetsch, F., Caille, I., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716 (1999).
26. Gutnick, M. J., Connors, B. W. & Ransom, B. R. Dye-coupling between glial cells in the guinea pig neocortical slice. *Brain Res.* **213**, 486–492 (1981).
27. Bergles, D. E. & Jahr, C. E. Synaptic activation of glutamate transporters in hippocampal astrocytes. *Neuron* **19**, 1297–1308 (1997).
28. Ge, W. P. & Duan, S. Persistent enhancement of neuron–glia signaling mediated by increased extracellular K⁺ accompanying long-term synaptic potentiation. *J. Neurophysiol.* **97**, 2564–2569 (2007).
29. Takano, T. *et al.* Astrocyte-mediated control of cerebral blood flow. *Nature Neurosci.* **9**, 260–267 (2006).
30. Rouach, N., Koulakoff, A., Abudara, V., Willecke, K. & Giaume, C. Astroglial metabolic networks sustain hippocampal synaptic transmission. *Science* **322**, 1551–1555 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. D. McCarthy and H. Zeng for providing us with the hGFAP-CreER and Ai14 transgenic mice, respectively; A. Sakaue-Sawano for the CAG-Fucci-Green transgenic mice; M. Stryker, Y. Xiang, X.-q. Wang, S. Barbel, J.-d. Chen, W. Zhou, F. Huang, and members of the Jan laboratory for discussion; G.-n. Li for help on electroporation; Y. Li for advice on retroviral experiments; and E. Unger, C. Guo, H. Yang, Q. Deng, J. Berg and X.-y. Li for reading the manuscript. W.-P.G. is a recipient of a Long-Term Fellowship of Human Frontier Science Program (HFSP) and National Institute of Neurological Disorders and Stroke (NINDS) Pathway to Independence Award. This work was supported by a NINDS K99/R00 award (1K99NS073735) to W.-P.G., a National Institute of Mental Health R37 grant (4R37MH065334) to L.Y.J., a National Institutes of Health (NIH) R01 grant (5R01MH084234) to Y.N.J., and grants from the NIH/National Institute on Aging P01 AG010435, MH090258, Jeffry M. and Barbara Picower Foundation (JBP) and McDonnell Foundation to F.H.G. L.Y.J. and Y.N.J. are Howard Hughes Medical Institute investigators.

Author Contributions W.-P.G. conceived the project, designed and performed the experiments and analysed the data. L.Y.J. and Y.N.J. supervised the work and helped to design the experiments. W.-P.G. and L.Y.J. wrote the manuscript. F.H.G. and A.M. provided MLV retrovirus and CAG-Fucci-Green transgenic mice, respectively. All authors reviewed and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.Y.J. (lily.jan@ucsf.edu).

METHODS

Animals and tamoxifen administration. The CAG-Fucci-Green transgenic line was from A.M.'s laboratory, the hGFAP-CreER line was from K. D. McCarthy's laboratory (UNC), and the NG2BacDsRed transgenic line was from A. Nishiyama's laboratory. Both NG2-CreBac³¹ and NG2-CreER were generated in Nishiyama's laboratory and bought from Jackson Laboratory. The Ai14 transgenic mice was from H. Zeng's laboratory. Tamoxifen inductions were as described¹⁸. For induction in hGFAP-CreER;Ai14 transgenic mice, an intraperitoneal or subcutaneous injection of tamoxifen (dissolved in a 1:10 mixture of ethanol and sunflower oil) at 3 mg per 40 g of body weight was administered once at the time indicated. All animals were treated in accordance with protocols approved by the Institutional Animal Care and Use Committee at UCSF.

In vivo electroporation. Newborn to 2-day-old pups (P0–2) were anaesthetized by hypothermia (about 4 min) and fixed to a support with a sticking plaster. GFP complementary DNAs were cloned into the chicken β -actin CMV promoter-driven expression vector pCAGGS. DNA solution (1–2 μ l) prepared at 2 mg ml⁻¹ in 10 mM Tris-HCl pH 8.0, with 0.04% trypan blue, was injected into the lateral ventricle with a pulled-out glass capillary (diameter 50–100 μ m)³². Animals were subjected to five electric stimuli of 50 V, each lasting 50 ms, at 950-ms intervals using a square-pulse electroporator BTX830.

Retroviral preparation and in vivo infection. pCAG-GFP-PRE contains replication-defective murine leukaemia virus (MLV)-based retroviral elements designed to carry and express enhanced GFP under CMV promoter and CAG promoter (modified chicken β -actin promoter with enhanced sequences from CMV) with control of the MLV long terminal repeat. We followed the detailed protocol for preparation from Gage's laboratory³³. In brief, three plasmids (pCAG-GFP-PRE, pCMV-gp and CMV-vsvg) were transfected to HEK 293T cells with Lipofectamine 2000. Viruses containing supernatants were harvested 2 days after transfection by centrifugation twice at 65,000g for 2 h (Discovery 90SE; Sorvall). Final virus titres were about 10⁶–10⁷ colony-forming units ml⁻¹ as measured by infecting HEK 293T cells. Viruses with the GFP reporter gene were injected (1 μ l) into either C57BL/6 wild-type or hGFAPCreER;Ai14 transgenic mice at P0–9. For *in vivo* infection, pups were anaesthetized with ice for 3–5 min, and the injection was performed as described³⁴. After injection, the pups were put back in a cage with a lamp to keep them warm. They were returned to their home cage when fully recovered.

Immunocytochemistry. Mice were perfused with 4% paraformaldehyde in 1 \times PBS. Brains were cut into sections 25–50 μ m thick with a cryostat (model CM3050S; Leica). Floating sections were permeabilized with 0.25% Triton X-100 in 1 \times PBS and then blocked for 2 h with 5% BSA and 3% normal goat serum with 0.25% Triton X-100 in 1 \times PBS. Primary antibodies for Ki67 (1:200 dilution, rabbit, monoclonal; Thermo Scientific), BLBP (1:1,000, rabbit, polyclonal; Invitrogen) or Laminin (1:500, rabbit, polyclonal; Sigma) were applied to sections alone or in combination and left to incubate for 24–48 h at 4 °C. Together with DAPI or Hoechst 33342 (1 μ g ml⁻¹; Invitrogen), secondary antibodies conjugated with Alexa488, 555, 568 or 633 (1:750) were applied for 2 h at room temperature (22–25 °C). To identify apoptotic astrocytes, sections were incubated for 15 min with 1 μ g ml⁻¹ propidium iodide after the treatment with 0.2 mg ml⁻¹ RNase (DNase-free) in 1 \times PBS for 30 min at 37 °C as described previously^{35,36}.

Slice preparation. Slices were prepared as described previously¹⁵. In brief, after decapitation, mouse brains were dissected rapidly and sliced with a vibratome (VT-1000S; Leica) in ice-cold oxygenated (95% O₂ and 5% CO₂) artificial cerebrospinal fluid solution (aCSF) containing (in mM): 119 NaCl, 2.5 KCl, 2.5 CaCl₂, 1.3 MgSO₄, 1 NaH₂PO₄, 26.2 NaHCO₃ and 11 glucose. Transverse slices (250 μ m in thickness) were then maintained in an incubation chamber for at least 1 h at room temperature before whole-cell recording, nuclear dye loading or time-lapse imaging.

Electrophysiology and live cell nuclear labelling. Whole-cell recordings from mouse brain slices were conducted with the aid of markers (GFP or Hoechst 33342) to identify infected cells or dividing cells. Astrocytes in hGFAP-GFP transgenic mice were identified by bright green fluorescence under the

microscope. For live nuclear labelling, slices were incubated with Hoechst 33342 (diluted to 2 μ g ml⁻¹ in aCSF) at room temperature for 30 min as described previously¹⁴. Recording pipettes were routinely filled with a solution containing (in mM): 125 potassium gluconate, 15 KCl, 10 HEPES, 3 MgATP, 0.3 Na-GTP, 5 Na-phosphocreatine and 0.2 EGTA (pH 7.2–7.4, 290–300 mosM). For glutamate transporter currents, pipette solution contained (in mM): 125 caesium gluconate, 5 CsCl, 10 HEPES, 3 MgATP, 0.3 Na-GTP, 0.2 EGTA and 5 Na-phosphocreatine (pH 7.2–7.4, 290–300 mosM). Membrane potential in voltage-clamp mode was held at –80 mV. Current pulses (20–60 μ A, 0.1 ms, 0.05 Hz) were delivered through extracellular bipolar electrodes placed roughly 200–300 μ m from the cells being recorded to induce transporter current.

Biocytin labelling. Glial cells were filled with 0.1% biocytin (e^N-biotinyl-L-lysine; Vector Lab) by means of a whole-cell recording electrode, as reported previously^{15,28}. Biocytin was dissolved in the recording pipette solution. Slices were fixed overnight with 4% paraformaldehyde at 4 °C before treatment for 2 h with blocking solution containing 5% BSA, 3% normal goat serum and 0.25% Triton X-100. Slices were then stained for 2 h with DyLight 549-conjugated streptavidin (1:1,000; Vector Lab). In Fig. 2o, DyLight 549 was added together with Alexa 633 (second antibodies against anti-Ki67) after washing out excess primary antibody against Ki67.

Confocal time-lapse imaging of acute brain slices. GFP⁺ cells at cortical slices from hGFAP-GFP transgenic mice (P3–5) were imaged on a Zeiss LSM510 two-photon confocal microscope equipped with objective 20 \times /0.5W and 63 \times /0.9W (Zeiss). Cells were scanned with xyz mode (four optical slices in z, with 8- μ m interval between slices). The frame interval was 4 min for 30–100 frames. Projection images were made from z-stacks that included all individual GFP⁺ cells. During imaging, slices were kept in a chamber with perfusion of aCSF (see above) at 32–34 °C.

Confocal time-lapse imaging in vivo. The pups (P3–6, hGFAP-CreER;Ai14 transgenic pups; Fig. 3l, m) were anaesthetized by hypothermia: 4–5 min in ice wrapped in a piece of cloth. A small fraction of skin (3 mm \times 3 mm) was removed over an area to be imaged. The pups then were returned to a box for 3–4 h until the incision site healed (no bleeding). A high-speed micro-drill was used to thin a circular area of skull, typically about 1 mm in diameter. The mouth of the pup was attached to a 1-ml pipette tip that was connected to a tube for inhalation. Pups were then immobilized with 4% agarose. Imaging was performed using a two-photon laser-scanning microscope based on a mode-locked laser system operating at 930 nm, equipped with one of the following objectives: 10 \times , 0.25 numerical aperture (NA); 20 \times , 0.8 NA collected emission more than 560 nm for tdTomato and 500–550 nm for mAG. Sometimes tdTomato was excited with a laser at 543 nm. Images were taken every 1.5 h for the first 3 h, and then the pups were put back in a box and allowed to move freely. Additional images were taken every 9–12 h for the following 18–24 h. During imaging, pups were fully anaesthetized with 2–4% isoflurane for 4–5 min. Cells under the pia were scanned with xyz mode (16 slices in z with 10- μ m interval between optical slices.). Shortly after an imaging session, isoflurane was turned off and oxygen was left on until the animal fully recovered. For Fucci-Green;hGFAP-CreER;Ai14 pups (Fig. 3i–k) we removed a small fraction of the skull (1 mm²) and images were taken within 1 h after surgery.

31. Zhu, X., Bergles, D. E. & Nishiyama, A. NG2 cells generate both oligodendrocytes and gray matter astrocytes. *Development* **135**, 145–157 (2008).
32. Li, G. *et al.* Regional distribution of cortical interneurons and development of inhibitory tone are regulated by Cxcl12/Cxcr4 signaling. *J. Neurosci.* **28**, 1085–1098 (2008).
33. Tashiro, A., Zhao, C. & Gage, F. H. Retrovirus-mediated single-cell gene knockout technique in adult newborn neurons *in vivo*. *Nature Protocols* **1**, 3049–3055 (2006).
34. Merkle, F. T., Mirzadeh, Z. & Alvarez-Buylla, A. Mosaic organization of neural stem cells in the adult brain. *Science* **317**, 381–384 (2007).
35. Barres, B. A. *et al.* Cell death and control of cell survival in the oligodendrocyte lineage. *Cell* **70**, 31–46 (1992).
36. Krueger, B. K., Burne, J. F. & Raff, M. C. Evidence for large-scale astrocyte death in the developing cerebellum. *J. Neurosci.* **15**, 3366–3374 (1995).

Optogenetic stimulation of a hippocampal engram activates fear memory recall

Xu Liu^{1*}, Steve Ramirez^{1*}, Petti T. Pang¹, Corey B. Puryear¹, Arvind Govindarajan¹, Karl Deisseroth² & Susumu Tonegawa¹

A specific memory is thought to be encoded by a sparse population of neurons^{1,2}. These neurons can be tagged during learning for subsequent identification³ and manipulation^{4–6}. Moreover, their ablation or inactivation results in reduced memory expression, suggesting their necessity in mnemonic processes. However, the question of sufficiency remains: it is unclear whether it is possible to elicit the behavioural output of a specific memory by directly activating a population of neurons that was active during learning. Here we show in mice that optogenetic reactivation of hippocampal neurons activated during fear conditioning is sufficient to induce freezing behaviour. We labelled a population of hippocampal dentate gyrus neurons activated during fear learning with channelrhodopsin-2 (ChR2)^{7,8} and later optically reactivated these neurons in a different context. The mice showed increased freezing only upon light stimulation, indicating light-induced fear memory recall. This freezing was not detected in non-fear-conditioned mice expressing ChR2 in a similar proportion of cells, nor in fear-conditioned mice with cells labelled by enhanced yellow fluorescent protein instead of ChR2. Finally, activation of cells labelled in a context not associated with fear did not evoke freezing in mice that were previously fear-conditioned in a different context, suggesting that light-induced fear memory recall is context-specific. Together, our findings indicate that activating a sparse but specific ensemble of hippocampal neurons that contribute to a memory engram is sufficient for the recall of that memory. Moreover, our experimental approach offers a general method of mapping cellular populations bearing memory engrams.

An important question in neuroscience is how a distinct memory is formed and stored in the brain. Recent studies indicate that defined populations of neurons correspond to a specific memory trace¹, suggesting a cellular correlate of a memory engram. Selective ablation or inhibition of such neuronal populations erased the fear memory response^{5,6}, indicating that these cells are necessary for fear memory expression. However, to prove that a cell population is the cellular basis of a specific fear memory engram it is necessary to conduct a mimicry experiment to show that direct activation of such a population is sufficient for inducing the associated behavioural output^{9,10}.

The hippocampus is thought to be critical in the formation of the contextual component of fear memories^{11–14}. Modelling¹⁵ and experimental^{16,17} studies have demonstrated an essential role of the dentate gyrus (DG) of the hippocampus in discriminating between similar contexts. Cellular studies of immediate early gene expression showed that sparse populations of DG granule cells (2–4%) are activated in a given context¹⁸. Moreover, although the same population of DG granule cells is activated repeatedly in the same environment, different environments¹⁹ or different tasks²⁰ activate different populations of DG granule cells. These lines of evidence point to the DG as an ideal target for the formation of contextual memory engrams that represent discrete environments.

To label and reactivate a subpopulation of DG cells active during the encoding of a memory, we targeted the DG of *c-fos*-tTA transgenic

mice³ with the AAV₉-TRE-ChR2-EYFP virus and an optical fibre implant (Fig. 1a). This approach directly couples the promoter of *c-fos*, an immediate early gene often used as a marker of recent neuronal activity²¹, to the tetracycline transactivator (tTA), a key component of the doxycycline (Dox) system for inducible expression of a gene of interest²². In our system, the presence of Dox inhibits *c-fos*-promoter-driven tTA from binding to its target tetracycline-responsive element (TRE) site, which in turn prevents it from driving ChR2-EYFP (enhanced yellow fluorescent protein) expression. In the absence of Dox, training-induced neuronal activity selectively labels active *c-Fos*-expressing DG neurons with ChR2-EYFP, which can then be reactivated by light stimulation during testing (Fig. 1b, c). We confirmed that our manipulation restricts the expression of ChR2-EYFP largely to the DG area of the hippocampus (Fig. 1d–g).

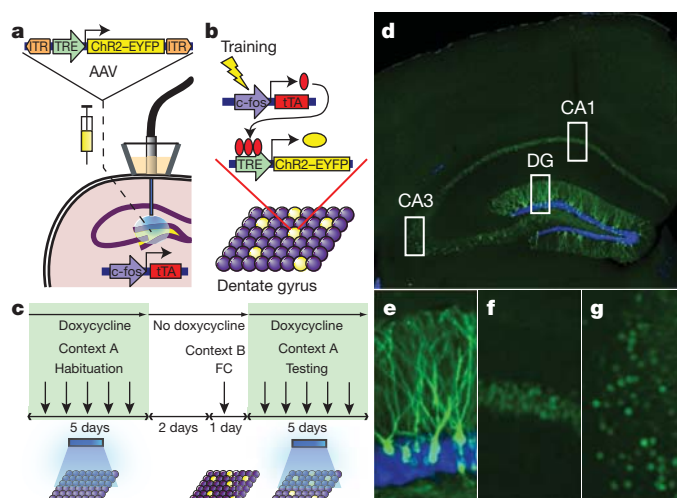


Figure 1 | Basic experimental protocols and selective labelling of DG cells by ChR2-EYFP. **a**, The *c-fos*-tTA mouse was injected with AAV₉-TRE-ChR2-EYFP and implanted with an optical fibre targeting the DG. **b**, When Dox, training induces the expression of tTA, which binds to TRE and drives the expression of ChR2-EYFP, labelling a subpopulation of activated cells (yellow) in the DG. **c**, Basic experimental scheme. Mice were habituated in context A with light stimulation while on Dox for 5 days, then taken off Dox for 2 days and fear-conditioned (FC) in context B. Mice were put back on Dox and tested for 5 days in context A with light stimulation. **d**, Representative image showing the expression of ChR2-EYFP in a mouse that was taken off Dox for 2 days and underwent FC training. **e–g**, An image of each rectangular area in **d** is magnified, showing the DG (**e**), CA1 (**f**) and CA3 (**g**). The green signal from ChR2-EYFP in the DG spreads throughout entire granule cells, including dendrites (**e**), whereas the green signal confined to the nuclei in CA1 and CA3 is due to a 2-h half-life EGFP (shEGFP) expression from the *c-fos*-shEGFP construct of the transgenic mouse (**f, g**). Blue is nuclear marker 4',6-diamidino-2-phenylindole (DAPI). Panel **d** is at $\times 10$ magnification and panels **e–g** are at $\times 50$ magnification.

¹RIKEN-MIT Center for Neural Circuit Genetics at the Picower Institute for Learning and Memory, Department of Biology and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Bioengineering and Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

First, to characterize the inducible and activity-dependent expression of Chr2-EYFP, we examined its expression timeline under various treatments (Fig. 2a–h). We observed a complete absence of Chr2-EYFP expression in DG neurons while mice were on Dox (Fig. 2a). Two days off Dox was sufficient to induce Chr2-EYFP expression in home-caged mice (Fig. 2b). The number of Chr2-EYFP-positive cells increased substantially in response to 2 days off Dox followed by fear conditioning (FC; Fig. 2c). We found that the vast majority of c-Fos-positive cells were also Chr2-EYFP positive (Supplementary Fig. 1), confirming that activity-dependent labelling with Chr2-EYFP recapitulated the induction of endogenous c-Fos. A similar increase in Chr2-EYFP expression was seen in a group of mice that were exposed to the same context and tone as the FC group but had no shocks delivered (NS; Fig. 2d). Chr2-EYFP expression lasted at least 5 days (Fig. 2e) and was gone by 30 days (Fig. 2f). Kainic acid-induced seizures resulted in complete labelling of DG cells with Chr2-EYFP (Fig. 2g), indicating that the relatively sparse labelling in

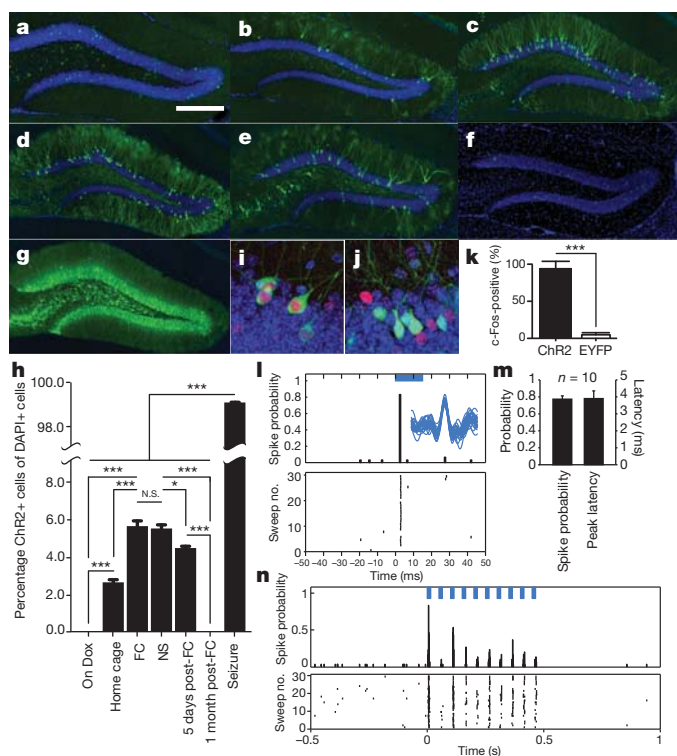


Figure 2 | Activity-dependent expression and stimulation of Chr2-EYFP. **a–g**, Representative images of the DG from c-fos-tTA mice injected with AAV₉-TRE-ChR2-EYFP and killed after the following treatments: on Dox (**a**); off Dox for 2 days in home cage (**b**); same as **b** followed by FC (**c**); same as **c** except no shock was delivered (NS; **d**); same as **c**, 5 days after training (**e**); same as **c**, 30 days after training (**f**); same as **b** followed by kainic acid injection to induce seizure (**g**). Residual green signal in **a** and **f** is from nuclear-localized c-fos-shEGFP (see Fig. 1 legend). **h**, Percentage of Chr2-EYFP-positive cells after various treatments represented by **a–g** ($n = 5$ subjects each; $F_{6,28} = 94.43$, $*P < 0.05$; $***P < 0.001$). N.S., not significant. **i, j**, Representative DG cells after light stimulation in c-fos-tTA mice injected with AAV₉-TRE-ChR2-EYFP (**i**) or AAV₉-TRE-EYFP (**j**). **k**, Percentage of c-Fos-positive cells among Chr2-EYFP-positive cells or EYFP-positive cells after light stimulation ($n = 3$ subjects each; $***P < 0.001$). **l**, Light-evoked single unit activity of a DG neuron from a c-fos-tTA mouse injected with AAV₉-TRE-ChR2-EYFP. Peri-event histogram (top) and raster plot (bottom) show reliable and precisely time-locked spiking relative to the onset of 15 ms light pulses (blue bar). Inset shows an overlay of waveforms for all the spikes during light stimulation. **m**, Spike probability and peak latency for all the light-responsive cells ($n = 10$) recorded as in **l**. **n**, Multi-unit activity in the DG from a c-fos-tTA mouse injected with AAV₉-TRE-ChR2-EYFP in response to trains of 10 light pulses (15 ms; blue bars) at 20 Hz. Scale bar in **a**, 250 μ m. Panels **a–g** are at $\times 10$ magnification and panels **i, j** are at $\times 80$ magnification. Error bars show mean \pm s.e.m.

the FC or NS groups was not due to the low infection rate of the virus, but reflected the naturally low activity of DG neurons during the training sessions^{18,23}. Notably, NS and FC treatments resulted in similar proportions of Chr2-EYFP-positive cells (Fig. 2h). Chr2-EYFP expression after FC seemed to be restricted to the excitatory neurons, as no overlap was detected between Chr2-EYFP-positive neurons and GABA-positive inhibitory neurons (Supplementary Fig. 2).

We injected c-fos-tTA mice with either AAV₉-TRE-ChR2-EYFP or AAV₉-TRE-EYFP, subjected them to FC while off Dox, and then put them back on Dox to test for light-evoked responses from DG cells the following day. The mice were anaesthetized for *in vivo* recordings, and blue light pulses (473 nm, 0.1 Hz, 15 ms pulse duration) were delivered to the DG. Consistent with the sparse labelling of DG neurons (Fig. 2h), we identified only ten DG neurons that responded to light stimulation from nine c-fos-tTA mice injected with AAV₉-TRE-ChR2-EYFP (the Chr2 group). In these neurons, we detected a reliable increase of spike probability precisely time-locked to the onset of light pulses (Fig. 2l, m). These cells also showed robust responses to trains of 20 Hz light stimulation, with a slight decrease in spike probability over time that remained higher above baseline (Fig. 2n). We did not find any light-responsive cells in the ten c-fos-tTA mice injected with AAV₉-TRE-EYFP (the EYFP group; data not shown). Most of the Chr2-EYFP-positive cells in the Chr2 group of mice were also positive for endogenous c-Fos after optical stimulation, although not all c-Fos-positive cells expressed Chr2-EYFP. Very few neurons expressing EYFP in the EYFP group of mice were c-Fos positive (Fig. 2i–k and Supplementary Fig. 3). The proportion of c-Fos-positive cells in the downstream CA3 region was greater in the Chr2 group compared with the EYFP group after optical stimulation of DG neurons, and this number was comparable to the proportion of CA3 c-Fos-positive cells obtained by exposing a separate group of mice to the conditioned context after FC (Supplementary Fig. 4).

Next, we tested whether activating a population of DG neurons labelled by Chr2-EYFP during the encoding of a fear memory was sufficient for memory recall. The experimental group (Exp) consisted of c-fos-tTA mice unilaterally injected with AAV₉-TRE-ChR2-EYFP and implanted with an optical fibre targeting the DG (Fig. 1a). Mice were kept on Dox and underwent 5 days of habituation to record their basal level of freezing in one context (context A) during both light-off and light-on epochs. Next, they were taken off Dox and underwent FC in a distinct chamber (context B) in which a tone was paired with shock. The mice were then subjected to 5 days of testing with light-off and light-on epochs in context A while on Dox (Fig. 1c). During the habituation sessions, the Exp mice showed very little freezing during either light-off or light-on epochs. In contrast, after FC, freezing levels during light-on epochs were higher compared with light-off epochs, which indicated light-induced fear memory recall (Fig. 3a). Increased freezing during light-on epochs was observed over all 5 days of test sessions with no discernible day-dependent difference (Supplementary Fig. 5g). These data suggest that DG cells that express endogenous c-Fos during training, and therefore become labelled by Chr2-EYFP, define an active neural population that is sufficient for memory recall upon subsequent reactivation.

To rule out the possibility that post-training freezing by optical stimulation was due to the activation of DG cells unrelated to fear learning, we injected another group of mice (NS) with AAV₉-TRE-ChR2-EYFP and administered the same habituation, training, and test sessions as the Exp group, except that no shock was delivered during the training session. Despite the fact that a similar level of Chr2-EYFP expression was detected in the NS group compared with the Exp group, both in terms of proportion of cells labelled (Fig. 2h) and Chr2-EYFP fluorescence intensity per cell (Supplementary Fig. 6), light did not induce post-training freezing in the NS group (Fig. 3b). This indicates that the freezing observed in the Exp group requires optical activation of a specific subset of Chr2-EYFP-positive DG cells that are associated with FC and that activating a population of DG cells

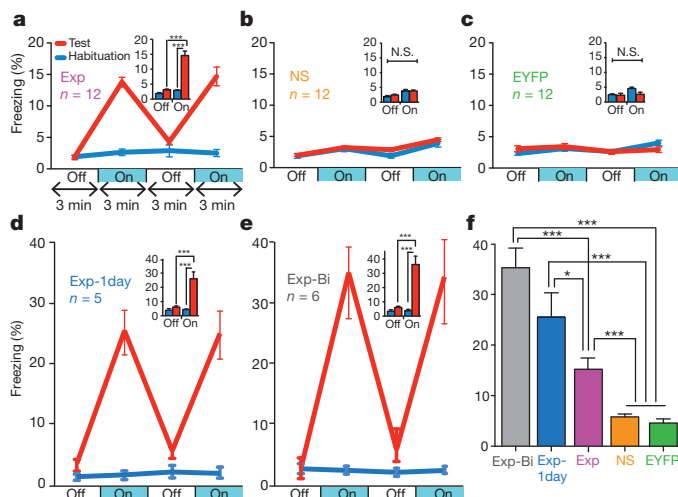


Figure 3 | Optical stimulation of engram-bearing cells induces post-training freezing. **a**, c-fos-tTA mice injected with AAV₉-TRE-ChR2-EYFP and trained with FC (Exp group) showed increased freezing during 3-min light-on epochs. Freezing for each epoch represents 5-day average (Supplementary Fig. 5a, g). Freezing levels for the two light-off and light-on epochs are further averaged in the inset ($n = 12$, $F_{1,22} = 37.98$, $***P < 0.001$). **b**, Mice trained similarly to the conditions in **a** but without foot shock (NS group) did not show increased light-induced freezing ($n = 12$). N.S., not significant. **c**, Mice injected with AAV₉-TRE-EYFP and trained with FC (EYFP group) did not show increased light-induced freezing ($n = 12$). **d**, Mice trained similarly to the conditions in **a** but kept off Dox for 1 day before FC training (Exp-1day group) showed greater freezing during test light-on epochs compared to Exp group ($n = 5$, $F_{1,8} = 38.26$, $***P < 0.001$). **e**, Mice trained similarly to the conditions in **a** but bilaterally injected with AAV₉-TRE-ChR2-EYFP and implanted with optical fibres (Exp-Bi group) showed even higher levels of freezing during test light-on epochs ($n = 6$, $F_{1,10} = 85.14$, $***P < 0.001$). **f**, Summary of freezing levels of the five groups during test light-on epochs ($F_{4,42} = 37.62$, $*P < 0.05$; $***P < 0.001$). Error bars show mean \pm s.e.m.

not associated with FC does not induce freezing. Yet another group of mice (EYFP) were injected with AAV₉-TRE-EYFP and underwent identical habituation, training, and testing sessions as the Exp group. The proportion of cells expressing EYFP was comparable to that seen in the Exp group expressing ChR2-EYFP (Supplementary Fig. 7). However, the EYFP group did not show increased post-training freezing (Fig. 3c). This result rules out the possibility that increased freezing in the Exp group was due to any non-specific effects of post-training optical stimulation.

The light-induced freezing levels of the Exp group were relatively low ($\sim 15\%$) compared with those typically reported from exposure to a conditioned context ($\sim 60\%$)³. One possibility is that light activation of background-activity-induced ChR2-EYFP (Fig. 2b) interfered with the expression of the specific fear memory. We confirmed that limiting the off-Dox period from 2 days to 1 day reduced the background expression of ChR2-EYFP by at least twofold (compare Supplementary Fig. 8a home cage with Fig. 2h home cage). A group of mice (Exp-1day) that went through the same design outlined in Fig. 1c but with this modification showed greater freezing levels ($\sim 25\%$) during the light-on epoch of test sessions compared to the Exp group (Fig. 3d, f). Another possible factor contributing to the modest light-induced freezing in the Exp group may be the limited number of cells optically stimulated. To test this possibility, we bilaterally injected a group of mice (Exp-Bi) with AAV₉-TRE-ChR2-EYFP and bilaterally implanted optical fibres targeting the DG, and then subjected these mice to the same scheme as that shown in Fig. 1c. During the light-on epochs of the test sessions, the Exp-Bi group exhibited levels of freezing ($\sim 35\%$) that were almost as high as those induced by the conditioned context (Fig. 3e, f, Supplementary Fig. 9 and Supplementary Movies).

We next examined whether the light-induced fear memory recall was context-specific. First, to test whether two different contexts activate

similar or distinct populations of DG cells, we took the mice off Dox for 2 days and then exposed them to a novel context (context C, an open field) to label the active DG cells with ChR2-EYFP. After being put back on Dox, the mice were fear-conditioned in a different context (context B) and killed 1.5 h later (Fig. 4a). The expression of ChR2-EYFP was used to identify cells previously activated in context C whereas endogenous c-Fos was used to identify cells recently activated in context B. Immunohistochemical analyses revealed a chance level of overlap between ChR2-EYFP-positive and c-Fos-positive cells, suggesting that two independent DG cell populations were recruited for the representation of the two distinct contexts (Fig. 4b–g). To test the context specificity of light-induced recall of a fear memory, we subjected a new group of mice (an open field fear-conditioned group; OF-FC) to habituation sessions in context A, followed by 2 days off Dox and exposure to context C to label neurons active in context C with ChR2-EYFP. Next, we put the mice back on Dox and performed FC in context B (Fig. 4h). These mice were then placed back in context A and tested for light-induced freezing. Light failed to evoke an increase in freezing responses (Fig. 4i). Similarly low levels of freezing were observed in another group of mice (FC-OFF) in which FC in context B while on Dox preceded exposure to context C while off Dox (Supplementary Fig. 10). Together, these results indicate that light reactivation of cells labelled in context C did not induce fear memory recall associated with context B.

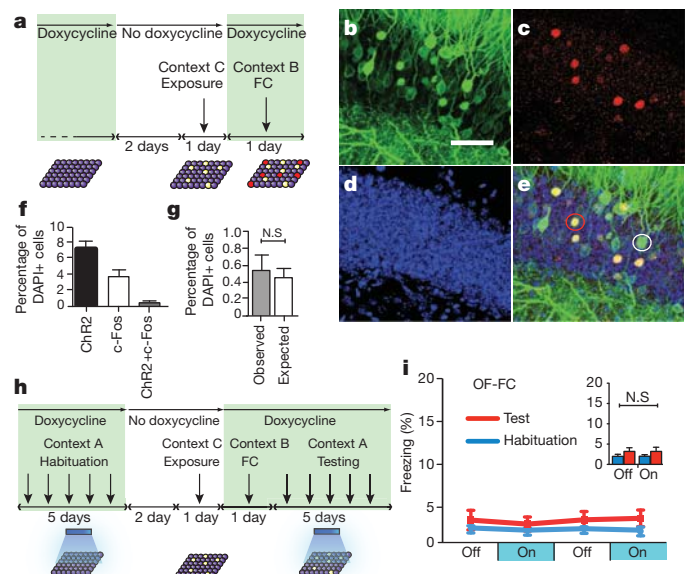


Figure 4 | Labelling and stimulation of independent DG cell populations. **a**, c-fos-tTA mice injected with AAV₉-TRE-ChR2-EYFP were taken off Dox and exposed to context C to label activated cells with ChR2-EYFP (yellow), then put back on Dox and trained with FC in context B to activate endogenous c-Fos (red). **b–e**, Representative images of DG from these mice are shown. **b**, ChR2-EYFP-labelled cells activated in context C. **c**, c-Fos-labelled cells activated in context B. **d**, Nuclear marker DAPI. **e**, Merge. The white and red circles show examples of ChR2-EYFP-positive and c-Fos-positive cells, respectively. The c-Fos-positive cells in **e** appear yellow because they express both endogenous c-Fos (red) and the nuclear-localized c-fos-shEGFP (green) (see Fig. 1 legend). **f**, Percentage of ChR2-EYFP-positive, endogenous c-Fos-positive, and double-positive cells among total cells (DAPI+) ($n = 5$). **g**, The observed percentage of double-positive cells is the same as what would be expected if the two cell populations were independent (that is, a product of the observed percentage of ChR2-EYFP single-positive and c-Fos single-positive cells). **h**, Behaviour setup for mice exposed to an open field in context C while off Dox and subsequently fear-conditioned in context B while on Dox (OF-FC). **i**, OF-FC mice ($n = 5$) do not show increased light-induced freezing. N.S., not significant. Panels **b–e** are at $\times 80$ magnification. Scale bar in **b**, 10 μ m. Error bars show mean \pm s.e.m.

We have shown that optical activation of hippocampal cells that were active during FC elicits freezing behaviour. To our knowledge, this is the first demonstration that directly activating a subset of cells involved in the formation of a memory is sufficient to induce the behavioural expression of that memory. Our results and previous studies that addressed the necessity of similarly sparse cell populations in the amygdala^{5,6} argue that defined cell populations can form a cellular basis for fear memory engrams. The memory engram that we selectively labelled and manipulated is probably contextual in nature, as previous studies have demonstrated that hippocampal interventions affect conditioned freezing responses to a context but not a tone^{12,13,24}. Indeed, recent findings show that optogenetic inhibition of the hippocampal CA1 region during training or testing inhibited the recall of a contextual fear memory, while leaving auditory-cued fear memory recall intact²⁵. However, we cannot completely rule out the possibility that the fear memory recalled in our experiments may have some tone memory component.

Our observation that freezing responses were elicited by optical stimulation in the experimental groups (Exp, Exp-1day and Exp-Bi), but not in the OF-FC or FC-OF group, strongly supports a dual memory engram hypothesis of contextual FC^{26–28}. In this hypothesis, hippocampal cells are recruited to form contextual memory engrams, but these contextual engrams alone do not represent a complete fear memory. For a fear memory to be formed, the information from the contextual memory engram must be transferred to the basolateral amygdala coincidentally with the information representing a foot shock. In the OF-FC or FC-OF scheme, two distinct contextual memory engrams were formed in the DG, which were represented by two distinct sets of DG cells. One of these two contextual engrams (the one for context B) was associated with the representation of the shock, but not the other engram (the one for context C). Because only the latter, but not the former, was labelled by ChR2, optical stimulation could not elicit fear memory expression.

Although we have demonstrated the sufficiency of a DG memory engram for the behavioural expression of a fear memory, we cannot conclude that this engram is necessary for behavioural recall. During contextual FC, it is likely that multiple contextual memory engrams are formed in a series of hippocampal regions. Each of these engrams may contribute to the formation of the complete fear memory in the BLA and may also be capable of reactivating it independently, as we observed in the case of the DG engrams. Because the hippocampus is not a linear feed-forward network but contains several parallel circuits, inhibiting the formation or activation of contextual engrams in one region may not necessarily block the expression of the fear memory. For instance, disruption of contextual memory engrams in the DG could be circumvented by CA1 engrams, which could be generated through the direct input from the entorhinal cortex and may be sufficient to activate the fear memory engram in the BLA. Indeed, we recently generated a mouse mutant, which permitted us to demonstrate that the DG input to the CA3 is dispensable in the formation and retrieval of contextual fear memory¹⁷.

The approach and methods described in this work will be a powerful tool for mapping multiple component engrams, each contributing to an overall memory. A multifaceted analysis of these engrams and their interplay will reveal the nature of the overall memory engram.

METHODS SUMMARY

Virus-mediated gene expression. The pAAV-TRE-ChR2-EYFP and pAAV-TRE-EYFP plasmids were constructed by standard methods and packed as AAV₉ viruses. The viruses were stereotactically injected into the DG (0.15 µl).

Immunohistochemistry. Mice were killed after various treatments and brain slices were prepared for immunohistochemistry and confocal microscopy. Coronal sections were immunostained for EYFP and/or c-Fos. All imaging and analyses were performed blind to the experimental conditions.

In vivo recording. An optrode consisting of a tungsten electrode glued to a 200 µm optical fibre coupled to a 473 nm laser was used for optical stimulation

and extracellular electrical recordings in head-fixed, isoflurane-anaesthetized mice.

Behavioural tests. Mice used for behavioural tests were injected with AAV₉ virus and implanted with an optical fibre targeting the DG. All mice were habituated in context A while on 40 mg kg⁻¹ Dox for 5 days for 12 min per day with light stimulation (473 nm, 20 Hz, 15 ms) during minutes 4–6 and 10–12. The groups of mice were taken off Dox for 1 (Exp-1day) or 2 days (Exp, Exp-Bi, EYFP and NS) and fear-conditioned in context B with a tone, with (Exp, Exp-1day, Exp-Bi and EYFP) or without (NS) shock. The OF-FC group was taken off Dox for 2 days and exposed to context C without shock, then fear-conditioned in context B while on Dox. The FC-OF group was fear-conditioned in context B while on Dox, then taken off Dox for 2 days and exposed to context C without shock. All groups were put back on Dox and tested in context A for 5 days in a manner similar to that used in the habituation sessions. All groups except the NS group were then put back to context B for a contextual fear probe trial 1 day after the last light stimulation, followed by a cued tone probe trial in context D the next day. Freezing levels were scored by experimenters blind to all treatment conditions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 November 2011; accepted 15 March 2012.

Published online 22 March 2012.

- Josselyn, S. A. Continuing the search for the engram: examining the mechanism of fear memories. *J. Psychiatry Neurosci.* **35**, 221–228 (2010).
- Silva, A. J. *et al.* Molecular and cellular approaches to memory allocation in neural circuits. *Science* **326**, 391–395 (2009).
- Reijmers, L. G., Perkins, B. L., Matsuo, N. & Mayford, M. Localization of a stable neural correlate of associative memory. *Science* **317**, 1230–1233 (2007).
- Han, J. H. *et al.* Neuronal competition and selection during memory formation. *Science* **316**, 457–460 (2007).
- Han, J. H. *et al.* Selective erasure of a fear memory. *Science* **323**, 1492–1496 (2009).
- Zhou, Y. *et al.* CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nature Neurosci.* **12**, 1438–1443 (2009).
- Boyden, E. S. *et al.* Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
- Martin, S. J. & Morris, R. G. New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus* **12**, 609–636 (2002).
- Gerber, B., Tanimoto, H. & Heisenberg, M. An engram found? Evaluating the evidence from fruit flies. *Curr. Opin. Neurobiol.* **14**, 737–744 (2004).
- Lever, C. *et al.* Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* **416**, 90–94 (2002).
- Phillips, R. G. & LeDoux, J. E. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav. Neurosci.* **106**, 274–285 (1992).
- Kim, J. J. & Fanselow, M. S. Modality-specific retrograde amnesia of fear. *Science* **256**, 675–677 (1992).
- Ramamoorthi, K. *et al.* Npas4 regulates a transcriptional program in CA3 required for contextual memory formation. *Science* **334**, 1669–1675 (2011).
- Treves, A. & Rolls, E. T. Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4**, 374–391 (1994).
- McHugh, T. J. *et al.* Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. *Science* **317**, 94–99 (2007).
- Nakashiba, T. *et al.* Young dentate granule cells mediate pattern separation whereas old granule cells facilitate pattern completion. *Cell* doi:10.1016/j.cell.2012.01.046 (23 February 2012).
- Schmidt, B., Marrone, D. F. & Markus, E. J. Disambiguating the similar: The dentate gyrus and pattern separation. *Behav. Brain Res.* **226**, 56–65 (2012).
- Chawla, M. K. *et al.* Sparse, environmentally selective expression of Arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. *Hippocampus* **15**, 579–586 (2005).
- Satvat, E. *et al.* Changes in task demands alter the pattern of *zif268* expression in the dentate gyrus. *J. Neurosci.* **31**, 7163–7167 (2011).
- Kubik, S., Miyashita, T. & Guzowski, J. F. Using immediate-early genes to map hippocampal subregional functions. *Learn. Mem.* **14**, 758–770 (2007).
- Shockett, P. E. & Schatz, D. G. Diverse strategies for tetracycline-regulated inducible gene expression. *Proc. Natl Acad. Sci. USA* **93**, 5173–5176 (1996).
- Leutgeb, J. K., Leutgeb, S., Moser, M. B. & Moser, E. I. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
- Lee, I. & Kesner, R. P. Differential contributions of dorsal hippocampal subregions to memory acquisition and retrieval in contextual fear-conditioning. *Hippocampus* **14**, 301–310 (2004).
- Goshen, I. *et al.* Dynamics of retrieval strategies for remote memories. *Cell* **147**, 678–689 (2011).
- Seidenbecher, T., Laxmi, T. R., Stork, O. & Pape, H. C. Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. *Science* **301**, 846–850 (2003).

27. Schafe, G. E., Doyere, V. & LeDoux, J. E. Tracking the fear engram: the lateral amygdala is an essential locus of fear memory storage. *J. Neurosci.* **25**, 10010–10014 (2005).
28. Rogan, M. T., Staubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Huang, G. Lin, M. Ragon and X. Zhou for help with the experiments, T. Ryan, A. Rivest, J. Young, R. Redondo and G. Dragoi for comments and discussions on the manuscript, and all the members of the Tonegawa laboratory for their support. This work is supported by National Institutes of Health grants R01-MH078821, P50-MH58880 to S.T. and RIKEN Brain Science Institute.

Author Contributions X.L., S.R., A.G. and S.T. contributed to the study design. X.L., S.R. and P.T.P. contributed to the data collection and interpretation. X.L. cloned all constructs. X.L. and S.R. conducted the surgeries and the behaviour experiments. S.R. conducted the expression timeline experiments. P.T.P. conducted the electrophysiology experiments. C.B.P. contributed to the setup of the electrophysiology apparatus and wrote the Matlab software to analyse the data. K.D. provided the original ChR2 construct. X.L., S.R. and S.T. wrote the paper. All authors discussed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.T. (tonegawa@mit.edu).

METHODS

Subjects. The *c-fos*-tTA mice were generated from TetTag mice by breeding them with C57BL/6J mice and selecting those carrying only the *c-fos*-tTA transgene and not the bi-cistronic tetO promoter driving tau-LacZ and tTA^{H100Y} transgenes. These mice also contained a separate transgene consisting of a *c-fos* promoter driving the expression of nuclear-localized 2-h half-life EGFP (shEGFP), which is distinct from the whole-cell-localized Chr2-EYFP. Mice had food and water *ad libitum* and were socially housed until the beginning of the surgery. The mice were 8–12 weeks old at the time of surgery and had been raised on food containing 40 mg kg⁻¹ Dox for 4 weeks before surgery. Mice were single housed post-surgery and throughout the rest of the experiments. All procedures relating to mouse care and treatment conformed to the institutional and National Institutes of Health guidelines.

Virus construct and packaging. The pAAV-TRE-ChR2-EYFP plasmid was constructed by cloning TRE-ChR2-EYFP into an AAV backbone using the SpeI restriction site at the 5' terminus and the blunt end at the 3' terminus of the insert. The pAAV-TRE-EYFP plasmid was constructed by removing the Chr2 fragment from the pAAV-TRE-ChR2-EYFP plasmid using NheI and AgeI restriction sites, blunting with T4 DNA polymerase, and self-ligation of the vector, which retained the ATG start codon of the EYFP gene from the Chr2-EYFP fusion gene. The recombinant AAV vectors were serotyped with AAV₉ coat proteins and packaged by the Gene Therapy Center and Vector Core at the University of Massachusetts Medical School. Viral titres were 1×10^{13} genome copy ml⁻¹ for AAV₉-TRE-ChR2-EYFP and 1.5×10^{13} genome copy ml⁻¹ for AAV₉-TRE-EYFP.

Stereotactic injection and optical fibre implant. All surgeries were performed under stereotaxic guidance. Mice were anaesthetized using 500 mg kg⁻¹ avertin. The virus was injected using a glass micropipette attached to a 10 µl Hamilton microsyringe (701LT; Hamilton) through a microelectrode holder (MPH6S; WPI) filled with mineral oil. A microsyringe pump (UMP3; WPI) and its controller (Micro4; WPI) were used to control the speed of the injection. The needle was slowly lowered to the target site and remained for 10 min before the beginning of the injection. Mice for timeline studies and head-fixed electrophysiology recordings were injected bilaterally (−2.2 mm anteroposterior (AP); ±1.3 mm mediolateral (ML); −2.0 mm dorsoventral (DV)²⁹) with 0.15 µl AAV₉ virus at a rate of 0.1 µl min⁻¹. After the injection the needle stayed for five additional minutes before it was slowly withdrawn. The mice used for behaviour tests were unilaterally or bilaterally injected with the virus same as described above. After withdrawing of the needle, a Doric patchcord optical fibre (200 µm core diameter; Doric Lenses) precisely cut to the optimal length was lowered above the injection site (−2.2 mm AP; ±1.3 mm ML; −1.6 mm DV). Three jewelry screws were screwed into the skull surrounding the implant site of each hemisphere to provide extra anchor points. A layer of adhesive cement (C&B Metabond) was applied followed with dental cement (Teets cold cure; A-M Systems) to secure the optical fibre implant. A cap made from the bottom part of a 15 ml Falcon tube (for unilateral implant) or the top part of an Eppendorf tube (for bilateral implant) was inserted to protect the implant and the incision was closed with sutures. Mice were given 1.5 mg kg⁻¹ metacam as analgesic and remained on a heating pad until fully recovered from anaesthesia. Mice were allowed to recover for 2 weeks before all subsequent experiments. All fibre placements (Supplementary Fig. 11) and viral injection sites were verified histologically. As criteria we only included mice with Chr2-EYFP expression limited to the DG, which led to the exclusion of two mice throughout the study.

Chr2-EYFP and EYFP expression timeline. Fourteen days after surgery, subjects were either kept on Dox and immediately killed or taken off Dox for 1 or 2 days. The mice from the latter two groups were either killed with no further treatments (home cage), or underwent FC or NS protocols as described in the behaviour section below. After each treatment, mice were killed 1.5 h, 24 h, 5 days or 30 days later, as described in the main text, and underwent immunohistochemistry procedures. For seizure experiments, mice were taken off Dox for 2 days and injected intraperitoneally with 20 mg kg⁻¹ kainic acid. The mice were killed 6 h after the first behavioural onset of seizure.

In vivo recording. Mice were anaesthetized by isoflurane inhalation and placed in the stereotactic system with anaesthesia maintained with 0.5–1% isoflurane throughout the recording. An optrode consisting of a tungsten electrode (1 MΩ) glued to an optical fibre (200 µm core diameter; Doric Lenses), with the tip of the electrode extending beyond the tip of the fibre by 500 µm was used for simultaneous optical stimulation and extracellular recordings. The optrode was lowered to the dentate gyrus (−2.2 mm AP; 1.3 mm ML; −2.0 mm DV) using a hydraulic micromanipulator (Model 640; David Kopf Instruments). The optical fibre was connected to a 200 mW 473 nm laser (MBL F473; Opto Engine) and controlled by a function generator (33220A; Agilent Technologies). The power intensity of light emitted from the optrode was calibrated to about 9 mW, which was consistent with the power intensity used in the behavioural assays. To identify Chr2-labelled cells, light pulses of 15 ms were delivered at 0.1 Hz at the recording sites approximately

every 5–10 µm throughout the DG. After light responsive cells were detected, two types of light stimuli were tested: 15 ms light pulse every 10 s and a train of ten 15 ms light pulses at 20 Hz every 10 s. Unit activity was band-pass filtered (500 Hz–5 KHz) and acquired with an Axon Digidata 1440A acquisition system running Clampex 10.2 software. Data were analysed with custom software written in Matlab. After the recording, endogenous c-Fos expression was induced by delivering two epochs of 3-min light stimulation (9 mW, 20 Hz, 15 ms), separated by 3 min, to the DG, the same as in behavioural experiments (see below). Mice were killed and perfused 90 min later.

Immunohistochemistry. Mice were overdosed with avertin and perfused transcardially with cold PBS, followed by 4% paraformaldehyde (PFA) in PBS. Brains were extracted from the skulls and kept in 4% PFA at 4 °C overnight. Fifty-micrometre coronal slices were taken using a vibrotome and collected in cold PBS. For immunostaining, each slice was placed in PBST (PBS + 0.2% Triton X-100) with 5% normal goat serum for 1 h and then incubated with primary antibody at 4 °C for 24 h (rabbit anti-c-Fos 1:5,000, Calbiochem; rabbit anti-GABA 1:5,000, Abcam; chicken anti-GFP 1:500, Invitrogen). Slices then underwent three wash steps for 10 min each in PBST, followed by 1 h incubation with secondary antibody (1:200 AlexaFluor488 anti-chicken, Invitrogen; 1:200 AlexaFluor568 anti-rabbit, Invitrogen). Slices were then incubated for 15 min with DAPI (1:10,000) and underwent three more wash steps of 10 min each in PBST, followed by mounting and coverslipping on microscope slides.

Cell counting. To characterize the expression timeline of Chr2-EYFP and EYFP, the number of EYFP immunoreactive neurons in the DG were counted from six coronal slices (spaced 120 µm from each other) per mouse ($n = 5$ for Chr2 group, $n = 3$ for EYFP group). Coronal slices were taken from dorsal hippocampus centred on coordinates covered by our optical fibre implants (−1.94 mm to −2.46 mm AP; Supplementary Fig. 11). Confocal fluorescence images were acquired on a Leica TCS SP2 AOBs scanning laser microscope using a $\times 20/0.70$ NA oil immersion objective. The image analysis module Visiormorph DP within VIS (Visiopharm) calculated the number of Chr2-EYFP-positive or EYFP-positive cells per section by thresholding EYFP immunoreactivity above background levels and using DAPI staining to distinguish between nuclei. The analysis module also permitted isolation of only Chr2-EYFP-positive and EYFP-positive neurons by setting size and fluorescence thresholds to filter out nuclear-localized c-fos-shEGFP-positive cells. The cell body layer of DG granule cells was outlined as a region of interest (ROI) according to the DAPI signal in each slice. A similar protocol was followed for c-Fos-positive cell counts in DG and CA3, except a Cy3 filter was applied for the latter. For quantification comparisons, we used a one-way ANOVA followed by Tukey's multiple comparisons using $\alpha = 0.05$. Data were analysed using Microsoft Excel with the Statplus plug-in and Prism (GraphPad Software).

To analyse the overlap between c-Fos and Chr2-EYFP-expressing or EYFP-expressing cells, a z-stack method was used in conjunction with ImageJ³⁰ to montage ten optical stacks (1 µm each, step size 10 µm) taken under a $\times 20/0.70$ NA oil immersion objective. Separate GFP and Cy3 filtered images were digitally combined to produce composite images. Equal cutoff thresholds were applied to all captures to remove background autofluorescence. All imaging and analyses were performed blind to the experimental conditions. To quantify the expression levels of Chr2-EYFP per cell, an experimenter blind to each condition used ImageJ to calculate the fluorescence intensity signal as integrated density for ten randomly chosen DG cells per hippocampal slice ($n = 3$ slices per mouse, 5 mice per condition; Supplementary Fig. 6).

Behaviour assays. All the behaviour tests were conducted during the light cycle of the day. Four different contexts were used in the behaviour assays. Context A was a 30 × 25 × 33 cm conditioning chamber within a room with black walls, black curtains, and dim lighting. The chamber had a white plastic floor and was scented with 0.25% benzaldehyde. Context B was a 29 × 25 × 22 cm conditioning chamber within a second room with white walls and bright lighting. The chamber had a gridded floor and a triangular roof, and was scented with 1% acetic acid. Context C was a 41 × 41 × 31 cm unscented open field arena within a third room with white walls and intermediate lighting. Context D was a 29 × 25 × 22 cm conditioning chamber in the same room as context C. It had a white acrylic glass floor and was unscented. The experimental groups (Exp, Exp-1day and Exp-Bi) and EYFP control (EYFP) groups underwent exactly the same training protocol. During the habituation session, each mouse was introduced to context A daily for 5 days while on 40 mg kg⁻¹ Dox food. Each day the mouse was loaded into the chamber and the optical fibre implant was connected to a 473 nm laser (MBL F473; Opto Engine) controlled by a function generator (33220A; Agilent Technologies). The mouse was then allowed to explore the chamber for 12 min. The 12 min session was divided into four 3-min epochs, with the first and third epochs as the light-off epochs, and the second and fourth epochs as the light-on epochs. During the light-on epochs, the mouse received light stimulation (9 mW, 20 Hz, 15 ms) for the

entire 3-min duration. At the end of the 12 min, the mouse was immediately detached from the laser and returned to its home cage. Following the fifth habituation session, the mouse was kept on regular food without Dox for 1 (Exp-1day) or 2 (Exp, Exp-Bi and EYFP) days until the training session. On the training day the mice received three training trials separated by 3 h in their home cages. For each training trial, the mouse was kept in the conditioning chamber in context B for 500 s. A tone (20 s, 75 dB, 2,000 Hz) was turned on at 180 s, 260 s, 340 s and 420 s, each of which co-terminated with a foot shock (2 s, 0.75 mA). After the third training trial, the mouse was returned to its home cage and placed on food containing 1 g kg⁻¹ Dox overnight to rapidly turn off any additional ChR2-EYFP or EYFP expression. The test session started the next day and the mouse was switched back to food containing 40 mg kg⁻¹ Dox. The procedure for the 5-day test session was exactly the same as the habituation session in context A. The day after the last test session, the mouse was returned to the original context B and exposed to the chamber for 300 s for a retrieval session to assay contextual fear memory. The next day, the mouse was introduced to context D for cued fear memory retrieval. This session lasted for 780 s, with a tone (60 s, 75 dB, 2,000 Hz) turned on at 180 s, 420 s and 660 s. The no shock (NS) group went through the same habituation, training and test sessions as the Exp group, except that no foot shock was given during the training session. The open field fear-conditioned (OF-FC) group went through the same habituation sessions. After the fifth habituation session, Dox was removed from the mouse's diet for 2 days, followed by exposure to the open field arena in context C to allow for 10 min of active exploration. The mouse was subsequently returned to its home cage and placed on 1 g kg⁻¹ Dox food overnight. The following day, the mouse was fear-conditioned in context B in the same manner as described above. Test sessions were administered over 5 days in context A on 40 mg kg⁻¹ Dox food, and the OF-FC group also underwent context and tone probe trials after the 5 days of testing. The fear-conditioned open field (FC-OFF) group went through the

same habituation sessions. The day after the fifth habituation, the mouse was kept on 40 mg kg⁻¹ Dox food and went through the FC procedure in context B as described above. The mouse was placed off Dox for 2 days after FC. Then the mouse was exposed to the open field arena in context C and allowed to freely explore for 10 min, after which the mouse was returned to their home cage with 1 g kg⁻¹ Dox food overnight, followed by test sessions over 5 days in context A on 40 mg kg⁻¹ Dox food. Freezing behaviour for training, context, and tone probe trials was recorded with a digital camera and measured with FreezeFrame software (ActiMetrics). Light stimulation during the habituation and test sessions interfered with the motion detection of the program, and thus the freezing of these sessions was manually scored. Two experimenters scored each video independently in a double-blinded manner. The overall scores showed a <3% difference between the two experimenters and for simplicity only one set of scores from one experimenter was reported. The manual scoring and computer scoring of the same training videos gave similar freezing scores. For each group, within each session (habituation and test) and within each epoch (light-on and light-off), a one-way ANOVA with repeated measures followed by Tukey's multiple comparisons ($\alpha = 0.05$) revealed no difference over 5 days (Supplementary Fig. 5). We therefore averaged the freezing level over 5 days for each mouse. A two-way ANOVA with repeated measures followed by Tukey's multiple comparisons ($\alpha = 0.05$) revealed that only the experimental groups (Exp, Exp-1day and Exp-Bi) showed increases in averaged freezing levels for light-on epochs of test sessions compared to light-off epochs of test sessions and light-on epochs of habituation sessions (Fig. 3a, d, e).

29. Paxinos, G. & Franklin, K. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 2001).
30. Rasband, W. S. Image J. <http://imagej.nih.gov/ij/> (National Institutes of Health, 1997–2011).

DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation

Pierre Close^{1,2}, Philip East³, A. Barbara Dirac-Svejstrup¹, Holger Hartmann⁴, Mark Heron⁴, Sarah Maslen⁵, Alain Chariot², Johannes Söding⁴, Mark Skehel⁵ & Jesper Q. Svejstrup¹

Alternative messenger RNA splicing is the main reason that vast mammalian proteomic complexity can be achieved with a limited number of genes. Splicing is physically and functionally coupled to transcription, and is greatly affected by the rate of transcript elongation^{1–3}. As the nascent pre-mRNA emerges from transcribing RNA polymerase II (RNAPII), it is assembled into a messenger ribonucleoprotein (mRNP) particle; this is the functional form of the nascent pre-mRNA and determines the fate of the mature transcript⁴. However, factors that connect the transcribing polymerase with the mRNP particle and help to integrate transcript elongation with mRNA splicing remain unclear. Here we characterize the human interactome of chromatin-associated mRNP particles. This led us to identify deleted in breast cancer 1 (DBC1) and ZNF326 (which we call ZNF-protein interacting with nuclear mRNPs and DBC1 (ZIRD)) as subunits of a novel protein complex—named DBIRD—that binds directly to RNAPII. DBIRD regulates alternative splicing of a large set of exons embedded in (A + T)-rich DNA, and is present at the affected exons. RNA-interference-mediated DBIRD depletion results in region-specific decreases in transcript elongation, particularly across areas encompassing affected exons. Together, these data indicate that the DBIRD complex acts at the interface between mRNP particles and RNAPII, integrating transcript elongation with the regulation of alternative splicing.

The composition of mRNP particles has been the subject of a number of studies, using a variety of approaches (for example, ref. 5). There are likely to be different types of mRNP particles with distinct compositions and interaction partners. We sought to purify native mRNP particles and interacting proteins from the chromatin in which they are generated and in which they are active in co-transcriptional processes. As a starting point, we generated HEK293 cells expressing near-normal levels of Flag-tagged heterogeneous nuclear ribonucleoprotein (hnRNP) A1, an abundant hnRNP protein in human cells⁶. hnRNP A1 shuttles between the nucleus and the cytoplasm⁷, but at steady state it is mainly nuclear and concentrated in chromatin (Fig. 1a), from where it can be released by RNase A treatment (Fig. 1b, compare lanes 2 and 4). We used DNase I digestion and mild sonication to release mRNP particles from chromatin for purification. RNase inhibitors were present during the whole process, outlined in Fig. 1c. mRNP particles isolated by this approach are predominantly of nuclear (chromatin) origin (Supplementary Fig. 1). Native mRNP particles and their interacting partners were purified from chromatin isolated from ~10⁸ nuclei (Fig. 1d). Only two major bands (namely the added, proteinaceous RNase inhibitors) were detected after purification from control cells, whereas numerous proteins were detected in hnRNP A1-Flag elutions (Fig. 1d). These represent a heterogeneous mixture of core mRNP particle subunits and proteins interacting with such particles. Individual protein bands were excised and identified by mass spectrometric analysis (Fig. 1d; see also

Supplementary Fig. 2a). Most of the known 'core' mRNP proteins, such as the hnRNP proteins, were present in the purified fraction, confirming the biological relevance of this approach. Many other pre-mRNA processing proteins were also identified, including splicing factors, ATP-dependent RNA helicases and a substantial number of mRNA 3'-end

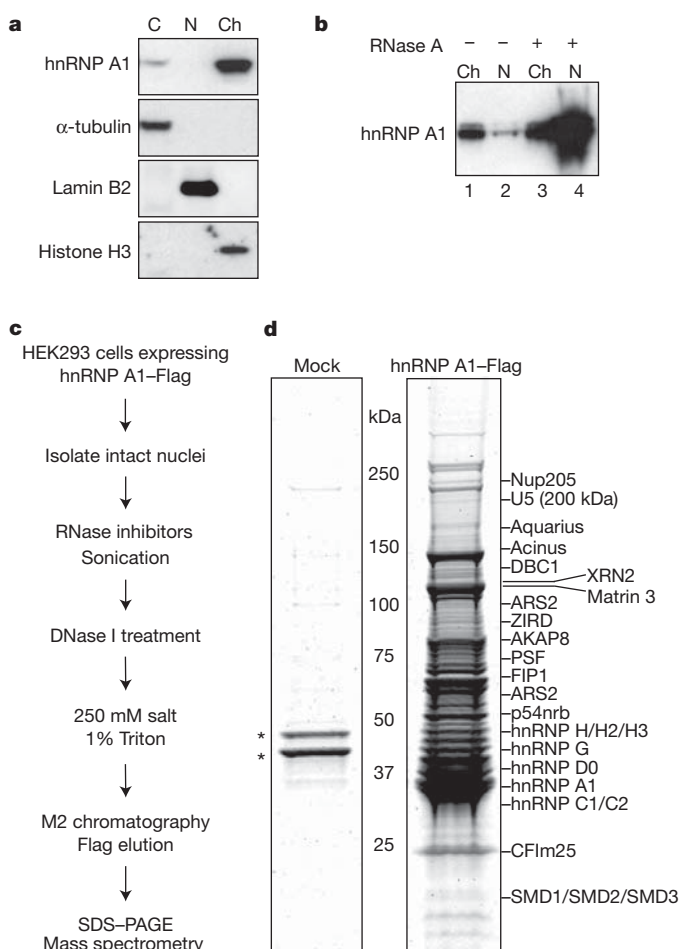


Figure 1 | Purification of nascent nuclear mRNP particles. **a**, Western blot analysis of cytoplasm (C), nucleoplasm (N) and chromatin (Ch), with α-tubulin, lamin B2 and histone H3 as controls for different fractions. **b**, Fractionation as in part **a**, but RNase A was added to the nuclear lysis buffer in the fractions indicated. **c**, Outline of the purification procedure. **d**, Equal fractions of the M2 chromatography eluates from control (Mock) and hnRNP A1-Flag separated by 4–12% SDS-PAGE and stained with SYPRO ruby. Asterisks mark RNase inhibitor proteins. Some of the identified proteins are indicated on the right.

¹Mechanisms of Transcription Laboratory, Cancer Research UK London Research Institute, Clare Hall Laboratories, South Mimms EN6 3LD, UK. ²Unit of Medical Chemistry, GIGA-Signal Transduction, GIGA-R, University of Liege, CHU, Sart-Tilman, 4000 Liege, Belgium. ³Bioinformatics & Biostatistics Group, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. ⁴Gene Center and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. ⁵Protein Analysis and Proteomics Laboratory, Clare Hall Laboratories, Cancer Research UK, London Research Institute, South Mimms EN6 3LD, UK.

processing and termination factors. Co-immunoprecipitation experiments confirmed the RNA-dependent interaction of some of these proteins with hnRNP A1–Flag (Supplementary Fig. 2b).

We next focused on two proteins that had not previously been connected to mRNP particles or mRNA processing. One of these, DBC1, is best known for its association with, and regulation of, the sirtuin-like deacetylase SIRT1 (refs 8, 9). We also investigated the uncharacterized zinc-finger-containing protein ZNF326. Stable cell lines expressing near-normal levels of Flag-tagged versions of these proteins were established, and co-immunoprecipitation experiments confirmed that both DBC1 and ZNF326 interact with mRNP particles in an RNA-dependent manner (Supplementary Fig. 3a–f). Furthermore, we discovered that ZNF326 and DBC1 associate directly, in an RNA-independent manner (Fig. 2a, e). For this reason, hereafter we refer to ZNF326 as ZNF-protein interacting with nuclear mRNPs and DBC1 (ZIRD).

We previously identified DBC1 as an RNAPII-interacting protein in another proteomic screen¹⁰, making it a particularly interesting candidate. Co-immunoprecipitation experiments confirmed that RNAPII associates with DBC1–Flag in an RNA-independent manner (Fig. 2b). Furthermore, ZIRD was detected in RNAPII (RPB3–Flag) purifications, and this interaction was also RNA-independent (Fig. 2c). In further support of a ZIRD–RNAPII interaction, ZIRD–Flag also co-immunoprecipitated RNAPII (Fig. 2d). In contrast, we failed to detect an interaction between hnRNP A1 and RNAPII under the same conditions (Fig. 2c, middle panel, and data not shown), although co-immunoprecipitation experiments after formaldehyde crosslinking indicated that, as expected, the proteins are in close proximity *in vivo* (Supplementary Fig. 4). Together, these results indicate that DBC1 and ZIRD are not part of the core mRNP particle, but that they might work at the interface between the mRNP particle and RNAPII.

Others reported that DBC1 interacts with SIRT1 (refs 8, 9). Although we observed that DBC1 co-precipitated SIRT1, endogenous ZIRD and ZIRD–Flag did not (Fig. 2e, and data not shown). SIRT1 is also absent from hnRNP A1-containing mRNP particles (Supplementary Fig. 3g). This indicates that ZIRD and DBC1 form a complex that lacks SIRT1. To characterize the ZIRD–DBC1 interaction further, ZIRD–Flag was purified. Size-exclusion chromatography of highly purified material showed that ZIRD–Flag and DBC1 are part of a salt-stable ~800-kDa complex (Fig. 2f) that also co-purified on MonoQ (data not shown). As expected, SIRT1 is not part of this protein complex (Fig. 2g, and data not shown). We named it the DBC1–ZIRD complex (DBIRD).

DBC1 and ZIRD interact with RNAPII in crude extracts (Fig. 2b–d). To investigate whether this interaction is direct, the DBIRD complex was characterized by gel filtration after mixing with an excess of RNAPII. In the absence of RNAPII, the DBIRD complex peaked in fractions 13–15 (Fig. 2f, upper two panels). However, when mixed with RNAPII (Fig. 2f, lower three panels), DBIRD complex elution shifted to earlier fractions, peaking in fraction 10 with a sub-fraction of RNAPII, whereas polymerase alone peaked in fractions 17–19 (~500 kDa), as expected. The DBIRD complex thus seems to form a bridging complex that interacts with both mRNP particles and RNAPII. Interestingly, DBIRD also interacted with mRNP particles lacking hnRNP A1 (Fig. 2h), pointing to a general bridging role.

To examine the role of the DBIRD complex in transcription-associated processes *in vivo*, we analysed the transcriptome of cells that had been depleted for DBC1 or ZIRD by RNA interference (RNAi) (Supplementary Fig. 5a). Total mRNA was hybridized to GeneChip Human Exon 1.0 ST arrays, on which the abundance of individual exons can be analysed independently. In the absence of ZIRD, a greater than 1.5-fold increase in exon inclusion was observed

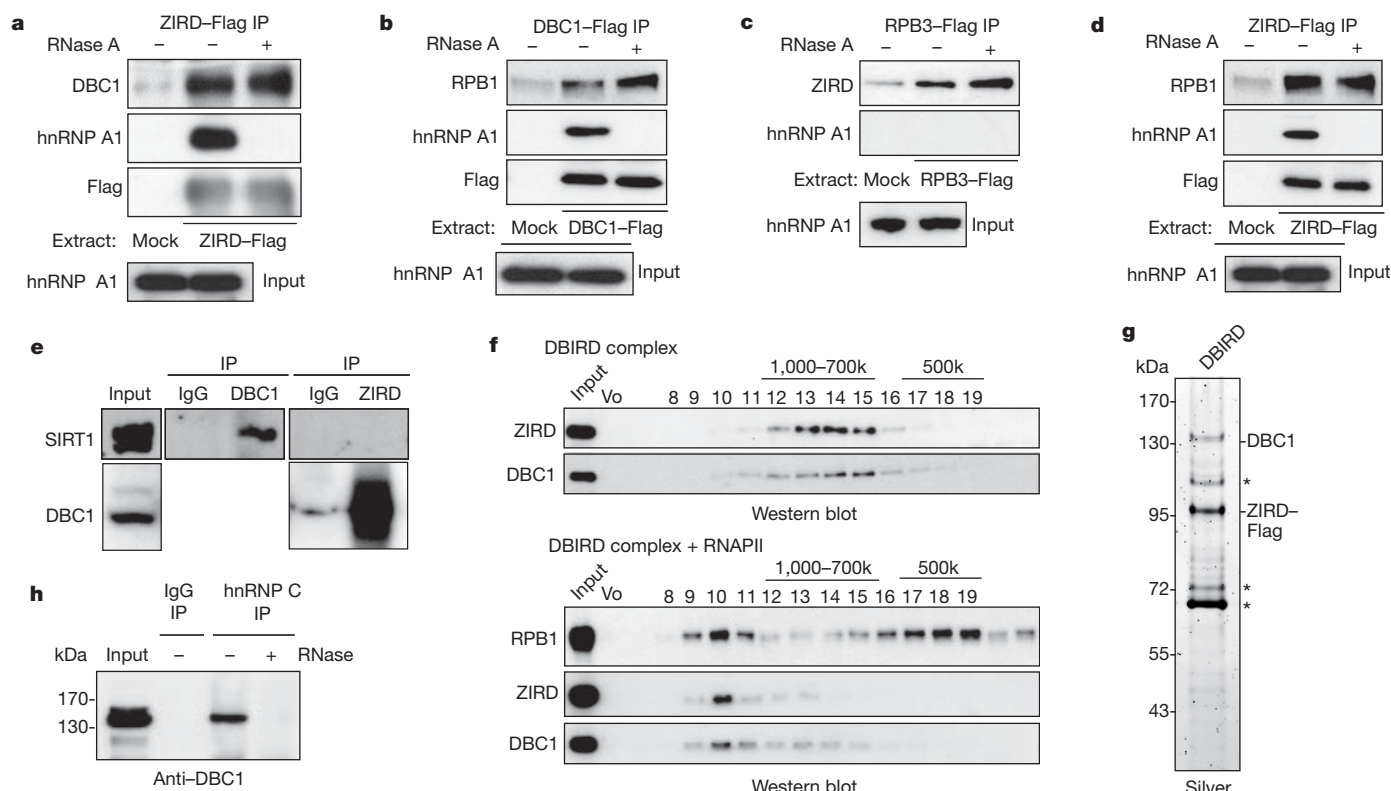


Figure 2 | DBC1 and ZIRD form a stable complex that binds RNAPII.

a, Western blot analysis of anti-Flag immunoprecipitates (IPs) from ZIRD–Flag cells. **b**, As in **a**, but for DBC1–Flag cells. **c**, As in **a**, but for RPB3–Flag cells. **d**, As in **a**, but detecting RPB1. **e**, Western blot analysis of anti-DBC1 and anti-ZIRD immunoprecipitates. IgG, immunoglobulin- γ . **f**, DBIRD analysed

by size-exclusion chromatography with (lower three panels) or without (upper two panels) RNAPII in approximately fivefold molar excess. Vo, void volume fraction. **g**, Silver stain of DBIRD. Asterisks indicate DBC1 and ZIRD degradation products. **h**, Western blot analysis of hnRNP-C-containing mRNP particles from mouse CB3 cells lacking hnRNP A1 (ref. 20).

in more than 2,800 situations, whereas exon exclusion was observed in only 390 cases (Supplementary Table 1a). The absence of DBC1 led to increased inclusion of an exon in 796 cases (Supplementary Table 1b) and, notably, most of these events were also on the list of ZIRD-dependent exon inclusions (567 out of 796 (71%); P -value for shared exons = 6.705×10^{-261} ; Supplementary Table 1c), which strongly supports a close functional relationship between the two factors and provides confidence in the genome-wide alternative splicing data sets. The effect was at the level of alternative splicing, as depletion of ZIRD or DBC1 only affected the expression of a very small number of genes (Supplementary Fig. 6).

A full list of inclusion events observed in both DBC1- and ZIRD-depleted cells is in Supplementary Table 1c. Sample results were confirmed by quantitative PCR with reverse transcription (RT-PCR) (Supplementary Fig. 7). To investigate whether DBIRD was present at affected exons, we performed RNA immunoprecipitation experiments¹¹. DBC1 and ZIRD bound the relevant exon in mRNAs from seven tested genes, whereas other regions (or control transfer RNA) were not detected or detected to a much lower extent (Fig. 3a, b and Supplementary Fig. 8). Interestingly, some exons of the β -actin gene (whose splicing was unaffected by DBIRD depletion) had considerable levels of DBIRD complex as well (Supplementary Fig. 8), indicating that the interaction of DBIRD with mRNA is not invariably correlated with DBIRD-dependent splicing changes.

To investigate the mechanism underlying exon inclusion, we first searched for sequence motifs in the DNA encompassing the included exons, but failed to uncover motifs other than those known to typify splice junctions. We then looked for nucleotide patterns that might be over-represented in the sequences surrounding the included exons by counting how often each of the 1,024 possible 5-base oligonucleotides occurred. Interestingly, (A + T)-rich 5-base oligonucleotides were markedly enriched around included exons (Fig. 3c). The frequencies of the four nucleotides in the regions around the splice sites were also analysed. A and T were strongly over-represented around the splice sites of DBIRD-affected exons, as well as across the exons themselves (Fig. 3d). The observed difference in A + T content is sufficient to explain the over-representation of (A + T)-rich 5-base oligonucleotides (Supplementary Fig. 9).

The (A + T)-rich DNA surrounding the affected exons might influence fundamental aspects of transcription. Indeed, A- and

T-tracts are difficult for RNAPII to transcribe, as they constitute very efficient elongation pause sites *in vitro*^{12,13}. To investigate the effect of DBIRD on transcript elongation, we performed RNAPII chromatin-immunoprecipitation (ChIP) analysis after DBIRD knockdown. For a control, we also knocked down SIRT1 (Supplementary Fig. 5b). Remarkably, although overall transcription of *RAD50* and *SLC36A4* is not affected (see Supplementary Fig. 7), depletion of DBC1 or ZIRD (but not SIRT1) markedly affected RNAPII transcription distinctively in regions encompassing affected exons (Fig. 4 and Supplementary Fig. 11). Quantification of newly produced mRNA by bromo-UTP incorporation supported the idea that elongation rates were decreased in these regions (Supplementary Fig. 10). DBIRD depletion also affected RNAPII density at other genes whose splicing was exon-specifically affected, whereas little or no change in RNAPII density was observed at the unaffected β -actin control gene, even at exons that had an elevated DBIRD level (Supplementary Figs 11 and 12; compare to Supplementary Fig. 8).

Our data support the idea that the DBIRD complex functions at the interface between core mRNP particles and RNAPII, affecting local transcript elongation rates and alternative splicing at a subset of (A + T)-rich exon-intron junctions (Supplementary Fig. 13). Notably, several studies have shown that the rate of RNAPII elongation affects the efficiency of splicing, with slow elongation favouring exon inclusion^{1,3}. Therefore, one possible explanation for our data is that the DBIRD complex acts as an elongation factor that facilitates transcript elongation across (A + T)-rich regions, and thereby affects alternative splicing of exons in these regions. It has also been suggested that exons in the nascent pre-mRNA become tethered to the elongating transcription complex^{14,15}. Given that DBIRD binds both mRNPs and RNAPII, it might affect such tethering, and thereby affect splicing.

DBC1 has been implicated in tumorigenesis as a potential tumour suppressor, regulating apoptosis and cell survival¹⁶. Whether the role of DBC1 in the DBIRD complex and alternative splicing affects tumorigenesis is an interesting possibility, particularly in light of the recent finding that genes encoding components of the splicing machinery are often mutated in myelodysplastic syndromes and related disorders¹⁷. ZIRD has not previously been characterized in human cells, but its mouse homologue, ZAN75, is highly expressed in neuronal tissues¹⁸, suggesting that regulation of the DBIRD complex might contribute to

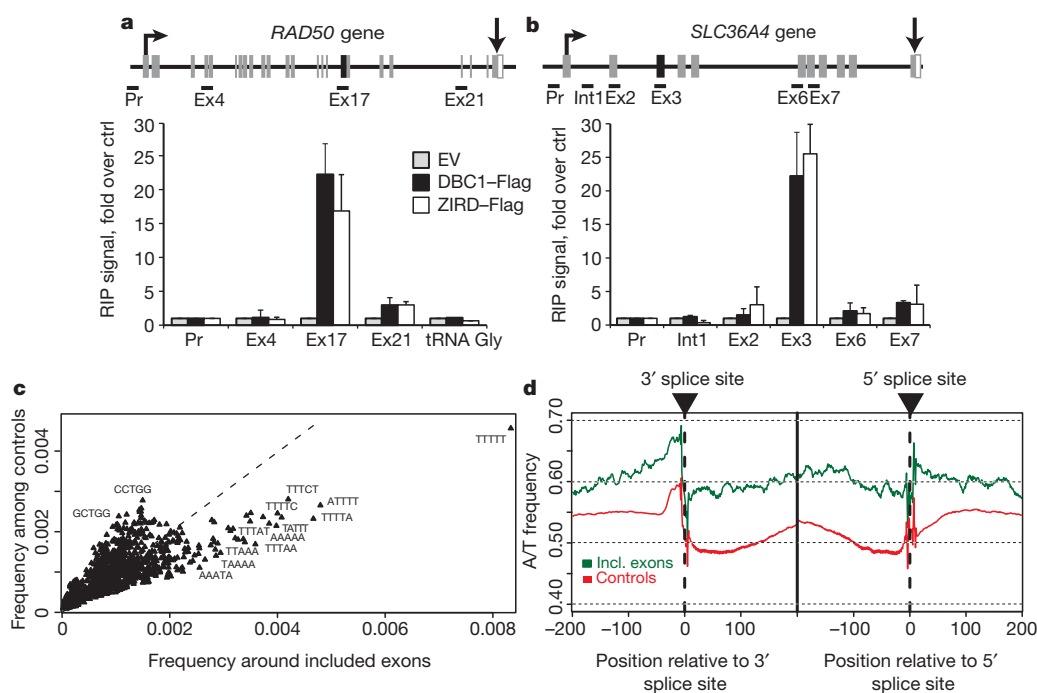


Figure 3 | DBIRD affects alternative splicing and is present at the affected exons. **a**, **b**, RNA immunoprecipitation (RIP) from crosslinked control (EV, ctrl), DBC1-Flag or ZIRD-Flag cells, analysed by quantitative PCR (qPCR). Control reactions lacking reverse transcriptase were always included (not shown). Error bars indicate standard deviations according to the Poisson statistic; $n = 3$. EV, empty vector; Ex, exon; Int, intron; Pr, promoter. Arrows indicate the start of transcription and the end of the coding region, respectively. **c**, Frequency of 5-base oligonucleotides in the regions around splice sites of affected (x axis) versus unaffected (y axis) exons. Diagonal line marks equal frequencies in the positive and negative set. **d**, Frequency of A or T upstream and downstream from splice sites of included exons (green) and unaffected control exons (red). The solid black line marks the meeting point of upstream and downstream data sets.

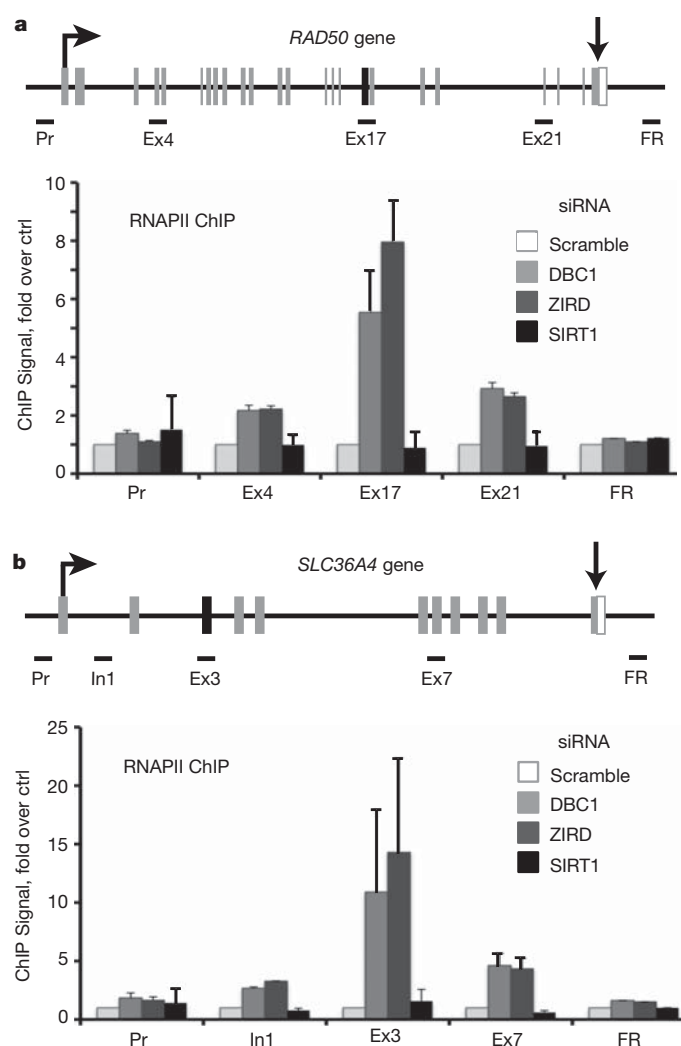


Figure 4 | DBC1 and ZIRD link exon skipping to RNAPII transcription. **a**, *RAD50* gene and qPCR primers (upper panel). RNAPII ChIP using cells transfected with control (scramble), *DBC1*, *ZIRD* or *SIRT1* Stealth siRNAs (lower panel). ChIP signals were normalized with inputs. Signals in control cells were set to 1 at each position, and values obtained from factor-depleted cells were expressed relative to these signals. Errors bars denote standard deviation; $n = 3$. **b**, Same as in **a**, but for *SLC36A4*. Supplementary Fig. 12 shows the same data in a format in which gene positional information is maintained. FR, flanking region. Arrows indicate the start of transcription and the end of the coding region, respectively.

tissue-specific splicing. Other proteins with homology to ZIRD and DBC1 exist in the human genome, raising the possibility that other DBIRD-like complexes are specific for other sets of genes or exons, or are involved in other transcription-related nuclear events.

METHODS SUMMARY

Open reading frames encoding hnRNP A1, DBC1 and ZIRD were cloned into pIRESpuro (Clontech) with a carboxy-terminal Flag tag. HEK293 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) containing 10% FBS in 5% CO₂ at 37 °C. For proteomic analysis, nuclei were isolated from hnRNP A1–Flag cells. These were sonicated and treated with DNase I, and the sample was cleared by centrifugation and the supernatant was subjected to M2 agarose (Sigma) chromatography. Proteins were eluted with 3× Flag peptide and mass spectrometry was performed as has been described elsewhere¹⁰. DBIRD was purified by M2 agarose chromatography from nuclease-treated nuclear extract from cells expressing ZIRD–Flag. DBIRD was analysed by MonoQ, or size exclusion chromatography with or without an excess of RNAPII. Stealth short interfering RNAs (siRNAs) were double transfected in HEK293 cells using lipofectamine 2000

(Invitrogen). For microarray analysis, RNA was hybridized on Human Exon 1.0 ST arrays (Affymetrix) using standard techniques (bioinformatics analysis described in Methods). For assessment of exon abundance and transcript expression, quantitative RT–PCR was performed using primers against affected and unaffected exons. Primer details are given in Supplementary Table 2. RNA immunoprecipitation and ChIP assays were performed as described^{11,19}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 May 2011; accepted 7 February 2012.

Published online 25 March 2012.

- Kornblihtt, A. R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
- Perales, R. & Bentley, D. “Cotranscriptionality”: the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**, 178–191 (2009).
- Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell Biol.* **3**, 195–205 (2002).
- Chen, Y. I. *et al.* Proteomic analysis of *in vivo*-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* **35**, 3928–3944 (2007).
- Kiledjian, M., Burd, C. G., Görlich, M., Portman, D. S. & Dreyfuss, G. In *Frontiers in Molecular Biology* (eds Mattaj, J. & Nigam, A. K.) 127–149 (Oxford Univ. Press, 1994).
- Piñol-Roma, S. & Dreyfuss, G. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**, 730–732 (1992).
- Kim, J. E., Chen, J. & Lou, Z. DBC1 is a negative regulator of SIRT1. *Nature* **451**, 583–586 (2008).
- Zhao, W. *et al.* Negative regulation of the deacetylase SIRT1 by DBC1. *Nature* **451**, 587–590 (2008).
- Aygun, O., Svejstrup, J. & Liu, Y. A RECQ5–RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc. Natl Acad. Sci. USA* **105**, 8580–8584 (2008).
- Selth, L. A., Close, P. & Svejstrup, J. Q. Studying RNA-protein interactions *in vivo* by RNA immunoprecipitation. *Methods Mol. Biol.* **791**, 253–264 (2011).
- Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol. Cell* **38**, 202–210 (2010).
- Saeki, H. & Svejstrup, J. Q. Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell* **35**, 191–205 (2009).
- Dye, M. J., Gromak, N. & Proudfoot, N. J. Exon tethering in transcription by RNA polymerase II. *Mol. Cell* **21**, 849–859 (2006).
- Fong, N., Ohman, M. & Bentley, D. L. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nature Struct. Mol. Biol.* **16**, 916–922 (2009).
- Kim, J. E., Chen, J. & Lou, Z. p30 DBC1 is a potential regulator of tumorigenesis. *Cell Cycle* **8**, 2933–2936 (2009).
- Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
- Lee, J. Y. *et al.* Characterization of a zinc finger protein ZAN75: nuclear localization signal, transcriptional activator activity, and expression during neuronal differentiation of P19 cells. *DNA Cell Biol.* **19**, 227–234 (2000).
- Close, P. *et al.* Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol. Cell* **22**, 521–531 (2006).
- Ben-David, Y., Bani, M. R., Chabot, B., De Koven, A. & Bernstein, A. Retroviral insertions downstream of the heterogeneous nuclear ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. *Mol. Cell Biol.* **12**, 4449–4455 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from Cancer Research UK and the European Research council (ERC) (to J.Q.S.), by an European Molecular Biology Organization (EMBO) long-term fellowship and by the Fonds Leon Fredericq foundation (to P.C.). We thank the Molecular Biology Core Facility at the Paterson Institute (Manchester) and Cell Services at London Research Institute for analysis and help. We thank B. Chabot for CB3 cells, and members of the Svejstrup laboratory, P. Verrijzer and T. H. Jensen for comments on the manuscript. P.C. is a senior research assistant at the Belgian National Foundation for Scientific Research (FNRS).

Author Contributions P.C. and A.B.D.-S. performed experiments, and S.M. and M.S. did mass spectrometry analyses. P.E., H.H., M.H. and J.S. performed bioinformatic analyses. P.C. and J.Q.S. designed the study, analysed the experimental data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Gene-expression and splicing data sets have been deposited in the GEO database under accession code GSE35480. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.Q.S. (j.svejstrup@canon.org.uk).

METHODS

Plasmids and antibodies. Open reading frames encoding human hnRNP A1, DBC1 and ZIRD (ZNF326) were cloned into pIRESpuro (Clontech) with a Flag tag at the C terminus. Antibodies used were rabbit anti-Flag, mouse anti-Flag M2 and mouse anti-hnRNP C (Sigma); mouse anti-pCTD mAb 4H8 (Millipore); rabbit anti-lamin B2 (Acris); rabbit anti-ZNF326 and mouse anti-hnRNP A1 (Santa Cruz Biotechnology); and rabbit anti-DBC1 and rabbit anti-SIRT1 (Bethyl Laboratories).

Cell culture, stable-cell-line establishment and stealth siRNA transfection. HEK293 cells were grown in DMEM containing 10% FBS in 5% CO₂ at 37 °C. To generate HEK293 stably expressing a Flag-tagged protein, cells were transfected with the relevant pIRESpuro construct and selected in 1 µg ml⁻¹ puromycin (Sigma). Cells were maintained in selecting media for 3 weeks, and surviving cells were used for experiments after transgene expression was checked.

Stealth siRNAs were double transfected in HEK293 cells using lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Protein and RNA expression was checked 48 h after the second transfection. Stealth siRNA anti-ZIRD RNAi sequences were: 5'-CGGAGGUAGUUAUGGUGUCGAUUU-3' (sense); 5'-AAAUCGACCACCAUAACUACCUCCG-3' (antisense). Stealth siRNA anti-DBC1 RNAi sequences were: 5'-CCAUCUGUGACUCCUAGAAGUCCA-3' (sense); 5'-UGGAGUUCUAGGAAGUCACAGAUGG-3' (antisense). For stealth siRNA anti-SIRT1 we used validated stealth siRNA (Invitrogen; oligo ID VHS0609). For control siRNA we used Stealth siRNA negative control med GC (Invitrogen; 12935-300).

Immunopurification of native mRNPs. 10⁸ cells stably expressing hnRNP A1-Flag were lysed with cytoplasmic lysis buffer (10 mM Tris HCl (pH 7.9), 340 mM sucrose, 3 mM CaCl₂, 2 mM Mg(OAc)₂, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 0.5% tert-butyl-type NP-40, protease inhibitors and 1 µl ml⁻¹ RNasin Ribonuclease inhibitor (Promega)), and intact nuclei were pelleted by centrifugation at 3,500g for 15 min. Nuclei were washed with cytoplasmic lysis buffer without NP-40 and then resuspended in DNase I buffer (20 mM HEPES (pH 7.9), 10% glycerol, 1.5 mM MgCl₂, 1 mM DTT, protease inhibitors and 1 µl ml⁻¹ RNasin). After ten strokes in a Dounce homogenizer, nuclei were sonicated using Bioruptor (Diagenode) before DNase I (Sigma) was added to the buffer and incubated for 30 min at room temperature. Buffer was then adjusted to a final concentration of 250 mM KOAc and 1% Triton X-100. The sample was cleared by centrifugation at 20,000g for 30 min and the supernatant was collected. For negative control purification, the same extracts were prepared from the same amount of untagged cells.

The sample was then applied to M2 agarose beads (Sigma) and incubated for 4 h at 4 °C. After binding, beads were washed extensively with washing buffer (20 mM HEPES (pH 7.9), 250 mM KOAc, 1% Triton X-100, 10% glycerol, 3 mM EDTA, 1 mM DTT, protease inhibitors and 1 µl ml⁻¹ RNasin Ribonuclease inhibitor)¹⁰. Finally, proteins were eluted by using Flag elution buffer (20 mM HEPES (pH 7.9), 100 mM KOAc, 3 mM EDTA, 1 mM DTT, 200 µg ml⁻¹ 3× Flag peptide, protease inhibitors and 1 µl ml⁻¹ RNasin Ribonuclease inhibitor). Eluates were resolved by 4–12% bis-Tris-gradient SDS-PAGE and revealed by SYPRO Ruby staining (Invitrogen).

Mass spectrometric analysis. Protein samples were reduced, alkylated and digested with trypsin, using the Janus liquid handling system (PerkinElmer). The digests were subsequently analysed by liquid chromatography tandem mass spectrometry on an LTQ Orbitrap XL mass spectrometer (ThermoScientific). The resulting data were searched against a protein database (UniProt KB) using the Mascot search engine programme (Matrix Science)²¹. All data were analysed manually.

Purification of the DBC1-ZIRD complex. Nuclei from approximately 10⁹ ZIRD-Flag cells were isolated, washed as above and then sonicated (Bioruptor, 30-s on-off cycles, max intensity for 15 min). Nucleic acids were digested by adding 10,000 units per ml Benzonase (Novagen) and 30 µg ml⁻¹ RNase A (Sigma), and incubating at 4 °C for 1 h. The nuclear extract was then adjusted to a final concentration of 250 mM KOAc and the insoluble fraction was removed by centrifugation (20,000g for 30 min). The supernatant was used for Flag-M2 chromatography. After extensive washes (60 column volumes of 20 mM HEPES-KOH (pH 7.9), 0.5% Triton X-100, 10% glycerol and 250 mM KOAc), bound proteins were eluted in the same buffer as above but containing 0.5 mg ml⁻¹ 3× Flag peptide (and 100 mM KOAc). The complex was then dialysed into buffer A (20 mM Tris-HCl (pH 7.9), 10% glycerol and 0.01% NP-40) containing 100 mM NaCl, before MonoQ PC 1.6/5 (GE Healthcare) or size exclusion (Superose 6 PC 3.2/30; GE Healthcare) chromatography. Proteins were eluted from MonoQ by a salt gradient from 0.1 to 1 M NaCl in buffer A. For size-exclusion chromatography, samples were loaded in buffer B (20 mM HEPES-KOH, 0.01% NP-40, 10% glycerol and 250 mM KOAc) with or without pre-incubation on ice with a five- to tenfold molar excess of purified mammalian RNAPII, purified as described²². Fifty-micro-

litre fractions were collected, and sizes were estimated by running protein size markers (Biorad) in parallel.

Microarray analysis. RNAs were extracted using RNeasy Kit (Qiagen), DNase-I-treated on the column, labelled and hybridized on Human Exon 1.0 ST arrays (Affymetrix) using standard techniques. Three independent experiments were performed and used as real triplicate for data analysis. We processed core-probe-level signals using robust multiarray average (RMA) implemented in APT (apt-1.10.0, Affymetrix) to generate quantile-normalized probe-set and gene-level signal estimates. Probe set to transcript cluster metagrouping was obtained from Affymetrix. We removed control probe sets from further analysis. For the gene-level analysis, genes displaying a coefficient of variance of less than 0.05 were assumed to be uninformative and were removed. We determined transcriptional effects (DBC1 versus control, ZIRD versus control and DBC1 versus ZIRD) by linear model, moderating the *t*-statistics by empirical Bayes shrinkage. We selected differential genes using a 0.05 *P*-value threshold using a nested F method. The analysis was carried out using the limma package from Bioconductor version 2.3 (ref. 23).

To identify putative alternative splicing events we first filtered probe sets to reduce false positive events. We removed all probe sets that did not localize to unique loci in the genome (Affymetrix annotation). We removed all probe sets from transcripts identified as not expressed in a given condition, as it is not possible to determine alternative splicing events against an untranscribed background. We defined a transcript as not being expressed if less than half of its member probe sets had a detection *P* value (detection above background value (dabg)) of less than 0.05 across all replicates. We also removed probe sets displaying a dabg value of greater than 0.05 in all replicates in both the conditions being considered. Genes with less than three probe sets after filtering were also removed from the analysis. To identify pairwise alternative splicing events we fitted transcript-cluster-specific linear models to probe set signal estimates and tested for significant interactions between each probe set and gene-level signal estimates across pairwise conditions using a 0.01 *P*-value threshold and a nested F method. In cases in which multiple probe sets mapped to a single exon, only probe sets that had significant interactions were included in the results. Once again the analysis was carried out using the limma package from Bioconductor²³.

Computational analysis of sequences around affected exons. To investigate sequence features that might explain the differential enrichment of certain exons after knockdown of ZIRD or DBC1, we selected a set of 505 exons that were 1.5-fold enriched in both the ZIRD and the DBC1 knockdowns and which were not the first or the last exon (to avoid including 5' and 3' untranslated regions in the following analyses). As a negative control set, we selected a further 3,877 exons that were unaffected after depletion of ZIRD and that were not the first or the last exon in the gene. We then prepared a positive and a negative sequence set for both the 5' and 3' splice sites. For the 5' splice site, the positive set contained the regions from -200 nucleotides to +50 nucleotides around the 5' splice sites of the 505 enriched exons. The negative control set contained corresponding regions for the 3,877 unaffected exons. Similarly, the positive set for the 3' splice site contained the region from -50 nucleotides to +200 nucleotides around the 3' splice sites of the 505 enriched exons, and the negative set contained the corresponding regions of the 3,877 unaffected exons.

The *P* values in Supplementary Fig. 9 were obtained in the following way. Suppose *ab* to be a dinucleotide and *i* a position around a splice site. We calculated the frequency $P_{\text{exp}}(i, ab)$ of *ab* that we would expect given the frequency of *ab* in the negative set, $P_{\text{neg}}(i, ab)$, and given the frequencies in the positive (P_{pos}) and negative (P_{neg}) sets of mononucleotide *a* at position *i*, and *b* at position *i* + 1, respectively:

$$P_{\text{exp}}(i, ab) = P_{\text{neg}}(i, ab) \frac{P_{\text{pos}}(i, a)P_{\text{pos}}(i + 1, b)}{P_{\text{neg}}(i, a)P_{\text{neg}}(i + 1, b)}$$

The *P* values for the observed number of dinucleotides *ab* at position *i* were calculated by approximating the binomial distribution with a normal distribution:

$$P \text{ value} = \sum_{k=K}^N \binom{N}{k} p^k (1-p)^{N-k} = \frac{1}{2} \operatorname{erfc} \left(\frac{K - Np}{\sqrt{Np(1-p)}} \right)$$

where *N* is the number of sequences in the set of affected exons, *K* is the number of observed dinucleotides *ab* at position *i*, *Np* is the expected number of dinucleotides at *i*, $p = P_{\text{exp}}(i, ab)$, *k* is the summation value, and *erfc* is the complementary error function.

Quantitative RT-PCR. RNAs were extracted using RNeasy Kit (Qiagen) and DNase-I-treated on the column, and 1 µg of total RNA was retro-transcribed using random primers and a first-strand complementary DNA synthesis kit (Fermentas). Quantitative RT-PCR was performed using SYBR Green detection. Specific primers against the alternatively spliced exon and unaffected exons were designed to assess exon abundance and transcript expression. Sequences are available on request.

Chromatin immunoprecipitation. ChIP assays were performed as described¹⁹, using 4H8 antibody (Abcam) or IgG antibody as a negative control, and incubation for 1 h with protein G/Herring sperm DNA. The precipitated DNA fragments were analysed using real-time PCR with SYBR Green detection. Input DNA was analysed simultaneously and used for normalization.

RNA immunoprecipitation. RNA immunoprecipitation was performed essentially as described^{11,24}. Briefly, 10⁷ HEK293 cells were crosslinked in 1% PFA for 10 min, quenched with 0.125 M glycine for 5 min, washed twice in PBS and collected in lysis buffer (50 mM Tris-HCl (pH 8), 1% SDS, 10 mM EDTA, and 50 U per 500 µl RNasin and protease inhibitors). Samples were sonicated for 15 min at maximum power using Bioruptor (Diagenode) to obtain RNA fragments 200–600 bases long, cleared by centrifugation and diluted tenfold in dilution buffer (20 mM Tris-HCl (pH 8), 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 50 U ml⁻¹ RNasin and protease inhibitors). Samples were pre-cleared by incubation with 100 µl of IgG agarose beads (Sigma) for 2 h at 4 °C. Anti-Flag M2 Resin (Sigma) was added and samples were incubated for 4 h at 4 °C. Immunoprecipitates were washed three times in wash buffer 150 (20 mM Tris-HCl (pH 8), 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, RNasin (50 U ml⁻¹) and protease inhibitors), once in wash buffer 500 (same as above, but with 500 mM NaCl), once in LiCl (10 mM Tris-HCl (pH 8), 250 mM LiCl, 0.5% NP-40, 0.1% Deoxycholate, 1 mM EDTA, 50 U ml⁻¹ RNasin and protease inhibitors) and once in TE100 buffer (TE containing 100 mM NaCl, 50 U ml⁻¹ RNasin, and protease inhibitors). Immunocomplexes were eluted three times in 150 µl elution buffer (TE containing 100 mM NaCl, 200 µg ml⁻¹ 3× Flag peptide, 50 U per ml RNasin) and incubated for 30 min at 37 °C with 1 µl of proteinase K. Crosslinking was then reversed by adding 9 µl of 5 M NaCl and incubating the samples at 65 °C for 1 h. Nucleic acids were extracted with phenol chloroform, then precipitated with ethanol and any remaining DNA was eliminated with Turbo DNase (Ambion) treatment. RNAs

were then reverse transcribed using random primers and the cDNAs were used for subsequent PCR reactions using relevant primers.

Nuclear run-on analysis. Cells were rinsed in PBS, then in buffer A (20 mM Tris-HCl (pH 7.4), 5 mM MgCl₂, 0.5 mM EGTA, 25% glycerol and 1 mM phenylmethylsulphonyl fluoride) and permeabilized in buffer A containing 0.02% Triton X-100 for 3 min at room temperature. The nascent RNA labelling reaction was carried out in buffer A containing 2 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 0.2 mM BrUTP (Sigma) and 25 U ml⁻¹ RNasin (Promega) for 15 min at 37 °C. In control reactions, normal UTP was used instead of BrUTP. After BrU incorporation, cells were rinsed twice in PBS and total RNA from nuclei of both labelled and control samples was isolated using TriPure Isolation Reagent (Roche). BrdU antibody (2 µg) (Sigma; also recognizes BrU) was pre-incubated with 20 µl of Protein G magnetic beads (Invitrogen) per experimental condition. RNA was then heated at 80 °C for 10 min and incubated with the beads at room temperature for 1 h with gentle shaking. The beads were washed five times in PBS containing 0.1% polyvinylpyrrolidone and RNasin (20 U per 200 µl), the RNA bound to the beads was eluted and the contaminant DNA was eliminated with Turbo DNase (Ambion). RNA was then reverse transcribed using random primers and the cDNAs were used for subsequent PCR reactions using relevant primers.

21. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
22. Hu, X. *et al.* A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl Acad. Sci. USA* **103**, 9506–9511 (2006).
23. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
24. Gilbert, C. & Svejstrup, J. Q. RNA immunoprecipitation for determining RNA-protein associations *in vivo*. *Curr. Protoc. Mol. Biol.* **75**, 27.4.1–27.4.11 (2006).

Differential positioning of adherens junctions is associated with initiation of epithelial folding

Yu-Chiun Wang^{1,2}, Zia Khan^{3,4†}, Matthias Kaschube^{4†} & Eric F. Wieschaus^{1,2}

During tissue morphogenesis, simple epithelial sheets undergo folding to form complex structures. The prevailing model underlying epithelial folding involves cell shape changes driven by myosin-dependent apical constriction¹. Here we describe an alternative mechanism that requires differential positioning of adherens junctions controlled by modulation of epithelial apical–basal polarity. Using live embryo imaging, we show that before the initiation of dorsal transverse folds during *Drosophila* gastrulation, adherens junctions shift basally in the initiating cells, but maintain their original subapical positioning in the neighbouring cells. Junctional positioning in the dorsal epithelium depends on the polarity proteins Bazooka and Par-1. In particular, the basal shift that occurs in the initiating cells is associated with a progressive decrease in Par-1 levels. We show that uniform reduction of the activity of Bazooka or Par-1 results in uniform apical or lateral positioning of junctions and in each case dorsal fold initiation is abolished. In addition, an increase in the Bazooka/Par-1 ratio causes formation of ectopic dorsal folds. The basal shift of junctions not only alters the apical shape of the initiating cells, but also forces the lateral membrane of the adjacent cells to bend towards the initiating cells, thereby facilitating tissue deformation. Our data thus establish a direct link between modification of epithelial polarity and initiation of epithelial folding.

The anterior and posterior dorsal transverse folds, or the dorsal folds, are epithelial folds that form on the dorsal side of the gastrulating *Drosophila* embryo at stereotypical locations coincident with the second and fifth stripes of the Runt expression (Fig. 1a–f, Supplementary Movie 1 and Supplementary Fig. 1a). Whereas the anterior fold is eventually shallow and the posterior fold deep, the initial cell shape changes are similar in both and the underlying mechanisms appear to be cell-autonomous (Supplementary Movies 2, 3 and Supplementary Fig. 1b, c).

We monitored cell shape changes using two-photon laser scanning microscopy in live embryos that express a membrane marker conjugated with the green fluorescent protein (Resille–GFP, also known as P{PTT-un1}CG8668^{117–2}). Optical sectioning of embryos at the mid-sagittal plane reveals that two stripes of dorsal cells, each three to seven cells wide, narrow their apices and shorten cell length during early gastrulation, producing two clefts on the dorsal surface that represent the first morphological signs of dorsal fold formation (Supplementary Fig. 2a and Supplementary Movie 4, see also Fig. 4b for measurements of shortening). Cells that undergo apical narrowing retain dome-like apices (Supplementary Fig. 2b), contrasting with the flattened apical surface caused by apical constriction during *Drosophila* ventral furrow formation².

We sought to identify dynamic cellular processes that precede cell shape changes. Unlike the canonical mode of epithelial folding in which spatially restricted activation of the molecular motor, myosin II (encoded by *spaghetti squash*), drives localized apical constriction to initiate tissue deformation^{1,3}, the basal levels of apical myosin remain

low and constant across the dorsal epithelium throughout the course of dorsal fold initiation with infrequent bursts of myosin activity that do not differ between the initiating and neighbouring cells (Supplementary Movie 5 and Supplementary Fig. 3a, b). These results indicate that the initiation of dorsal fold formation is not associated with differential myosin contractility.

In contrast, E-Cadherin (encoded by *shotgun*), the core component of adherens junctions, shows a cell-type-specific change in its positioning: in the initiating cells, junctions shift basally from the subapical regions where they are originally assembled, whereas in the neighbouring cells junctions maintain their original subapical positioning (Fig. 1g, Supplementary Movies 6 and 7 and Supplementary Fig. 3c). Simultaneous imaging of E-Cadherin–GFP and Resille–GFP reveals that basal shift of junctions can be observed as early as 300 s before the onset of gastrulation during the last phase of cellularization, which precedes the apical narrowing and cell shortening that occur 100–200 s after the onset of gastrulation (Fig. 1h and Supplementary Movie 8). During this seven-minute interval, junctions in the initiating cells shift approximately 10 μm basally to lie at $34 \pm 5\%$ ($n = 18$) below the apical surface, whereas junctions in the neighbouring cells show only a slight shift ($\sim 3 \mu\text{m}$) to lie at $15 \pm 4\%$ ($n = 27$) below the apical surface (Supplementary Fig. 4). The basal shift of junctions in the initiating cells increases the asymmetry in the junctional positioning on the opposite sides of the neighbouring cells that immediately flank the initiating cells. The lateral membrane of these cells becomes increasingly curved, correlating with the increased junctional asymmetry (Supplementary Fig. 5).

If the apparent basal shift of E-Cadherin positioning reflects an actual movement or remodelling of the junctions, it should be associated with an increase in the volume and surface area above the junctions. To test this hypothesis, we measured the two-dimensional parameters of area and perimeter of the apical domain in the living embryos. As the junctions shift basally in the initiating cells, both of these parameters increase, consistent with a basal movement of the junctions within the cells (Supplementary Fig. 6). We corroborated these observations by developing computer software that reconstructs and quantifies three-dimensional cell shape in fixed embryos (Fig. 1i). As the cell length increases during the last phase of cellularization, the length, volume and surface area of the apical domain in the initiating cells all increase significantly more than they do in the neighbouring cells (Fig. 1j–l), indicating that the junctional shift is accompanied by an expansion of the apical domain and that mobility of the E-Cadherin complex underlies the apparent basal shift of the junctions.

Adherens junctions are positioned to the subapical regions of the polarized epithelial cells by the concerted action of the scaffolding protein Par-3 (encoded by *bazooka* in *Drosophila*), the atypical protein kinase C (aPKC) and the MARK family kinase Par-1: apically localized aPKC and basal-laterally localized Par-1 restrict Par-3 to the subapical regions, where it directs junctional assembly^{4–10}. We found that the

¹Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. ²Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA. ³Department of Computer Science, Princeton University, Princeton, New Jersey 08540, USA. ⁴Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. [†]Present addresses: Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA (Z.K.); Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Frankfurt am Main 60483, Germany (M.K.).

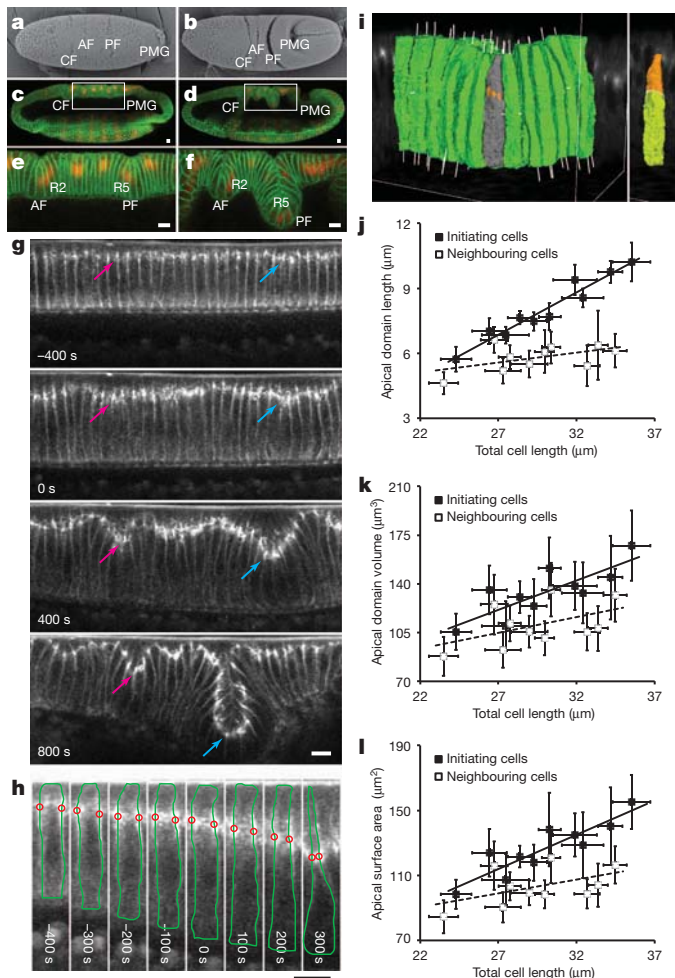


Figure 1 | Morphology and cellular dynamics during dorsal fold formation. **a, b**, Scanning electron micrographs of the dorsal surface in an early (**a**) and a late (**b**) *Drosophila* gastrula. **c–f**, Confocal mid-sagittal sections of Neurotactin (green) and Runt (red) immunofluorescence in an early (**c**) and a late (**d**) gastrula. **e, f**, A magnified view of the highlighted areas in **c** and **d**. AF, anterior fold; CF, cephalic furrow; PF, posterior fold; PMG, posterior midgut; R2, second stripe of Runt; R5, fifth stripe of Runt. **g**, Two-photon time-lapse mid-sagittal section of E-Cadherin-GFP. Arrows, junctions of initiating cells in the anterior (pink) and posterior (cyan) folds undergo basal shift. **h**, Two-photon time-lapse mid-sagittal section of E-Cadherin-GFP and Resille-GFP in a posterior fold-initiating cell with manual traces of cell outlines (green) and junctional position (red). Scale bars, 10 μm . **i**, Three-dimensional rendering of a posterior fold based on Neurotactin immunofluorescence with an initiating cell highlighted in grey (left panel, orange patches depict thresholded Bazooka staining). The Bazooka staining is used to subdivide the cell into the apical (orange) and basal (yellow) domains (right panel). **j–l**, Scatter plots of the average length, volume and surface area of apical domain against the average total cell length in a series of fixed late cellularizing embryos with solid and dashed trend lines for the initiating and neighbouring cells. Error bars indicate s.d.

levels of Bazooka and aPKC are not differentially regulated across the dorsal epithelium and thus do not account for the observed junctional shift (Fig. 2a, Supplementary Movie 9 and Supplementary Fig. 7). In contrast, live imaging of Par-1-GFP shows that the levels of Par-1 in the presumptive initiating cells, although initially similar ($\sim 95\%$) to those in the neighbouring cells before the onset of junctional shift, reduce progressively during the last phase of cellularization to reach approximately 80% of its levels in the neighbouring cells as gastrulation commences (Fig. 2b, c and Supplementary Movie 10, $n = 7$). This differential modulation of Par-1 levels seems to require the anterior–posterior patterning system (Supplementary Fig. 8 and Supplementary Movie 11). To ask whether the reduction in Par-1 levels in the initiating

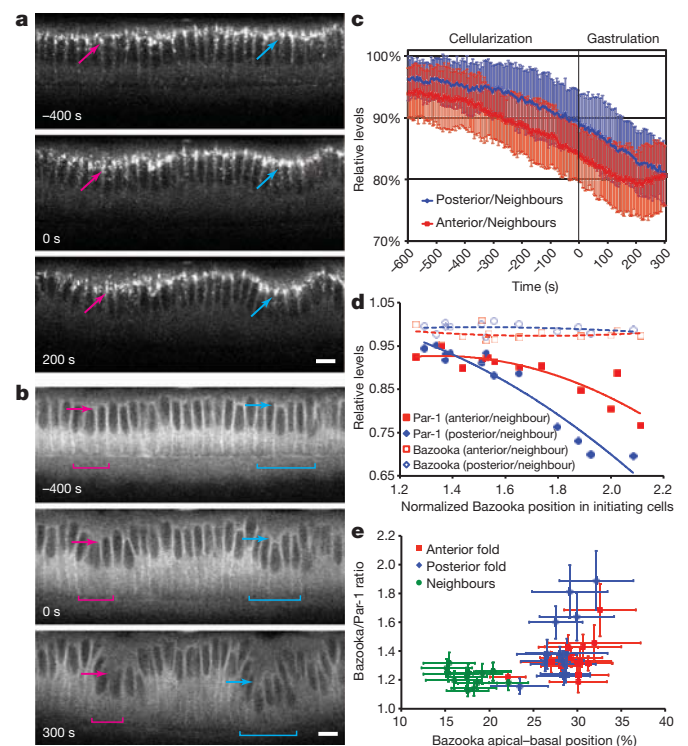


Figure 2 | The dynamics of Bazooka and Par-1 during dorsal fold initiation. **a, b**, Two-photon time-lapse mid-sagittal section of Bazooka-GFP (**a**) or Par-1-GFP (**b**). Arrows and brackets, anterior (pink) and posterior (cyan) fold-initiating cells. Scale bars, 10 μm . **c**, A time-course analysis of Par-1-GFP levels in the initiating cells relative to those in the neighbouring cells ($n = 7$). **d**, A scatter plot of the average Bazooka positioning in the initiating cells normalized by that in the neighbouring cells against the average levels of Bazooka or Par-1 relative to their respective levels in the neighbouring cells with the corresponding trend lines. **e**, A scatter plot of the average Bazooka positioning along the apical–basal axis against the average Bazooka/Par-1 ratio within individual cells. Error bars indicate s.d.

cells correlate temporally with the junctional shift, we quantified the levels of Par-1 in fixed embryos and determined the position of junctions using Bazooka staining. As Bazooka becomes more basally positioned in the initiating cells, their Par-1 levels also become lower, whereas the Bazooka levels remain constant (Fig. 2d). These analyses confirm our live imaging data and establish a correlation between the position of junctions and the ratio of Bazooka/Par-1 (Fig. 2e).

This correlation suggests that Par-1 downregulation allows Bazooka to gradually localize more basally, which in turn directs basal repositioning of junctions. To test this hypothesis, we altered the levels of Bazooka and Par-1 to investigate the function of junctional positioning during the formation of dorsal folds. Uniform reduction of Bazooka activity by RNA interference (RNAi) causes accumulation of E-Cadherin-GFP at the edges between apical and lateral surfaces, resulting in an extreme apical positioning of junctions across the epithelium (Fig. 3a and Supplementary Movie 13), similar to embryos produced by the germline clones of a strong loss-of-function allele of *bazooka* (Supplementary Fig. 9). Conversely, in *par-1* RNAi embryos, junctions are located in the lateral regions of all dorsal cells at an average position of $39 \pm 8\%$ below the surface, slightly more basal than the junctions in the initiating cells in the wild-type (Fig. 3b and Supplementary Movie 14, 30 cells from 3 embryos). Importantly, in both *bazooka* and *par-1* RNAi embryos, the junctional positioning is uniform across the entire dorsal epithelium and in each case, the initiation of dorsal folds is abolished despite the normal appearance of junction and epithelial structure (75% for *bazooka* RNAi, $n = 8$; 70% for *par-1* RNAi, $n = 10$). Thus, dorsal fold formation seems to

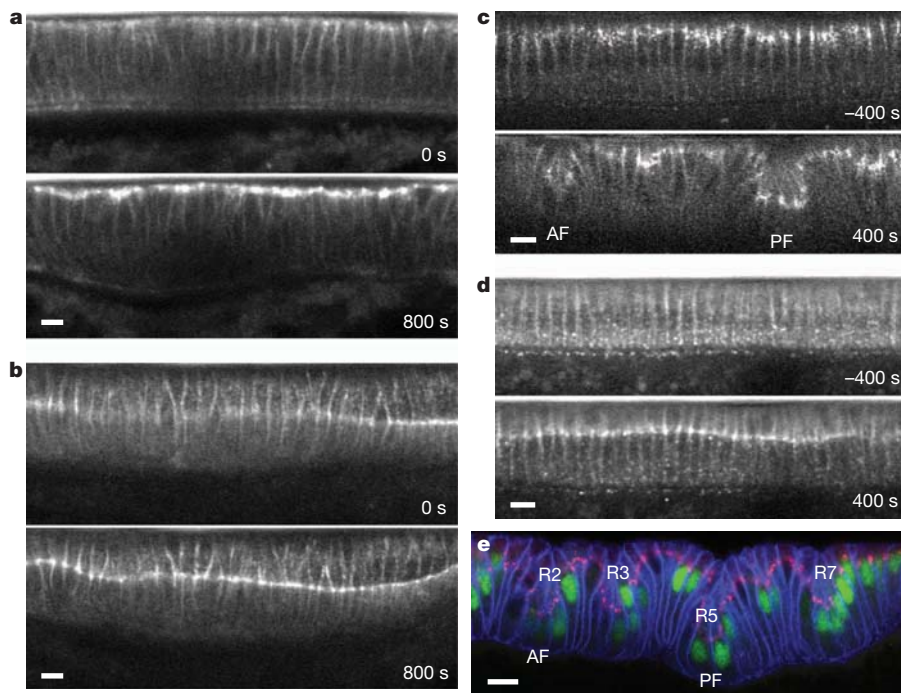


Figure 3 | Differential positioning of adherens junctions is necessary and overexpression of Bazooka can be sufficient for ectopic dorsal fold initiation. **a–d**, Two-photon time-lapse mid-sagittal section of E-Cadherin-GFP in a *bazooka* (**a**) or *par-1* (**b**) RNAi embryo or Bazooka^{S151A, S1085A}-GFP in an embryo in which the endogenous Bazooka is present (**c**) or downregulated

by *bazooka* 5'UTR RNAi (**d**). **e**, Confocal mid-sagittal section of Bazooka (red), Neurotactin (blue) and Runt (green) immunofluorescence in a Bazooka-GFP overexpression embryo. R2, R3, R5 and R7 denote the Runt stripes. AF, anterior fold; PF, posterior fold. Scale bars, 10 μ m.

require a differential positioning of junctions between the initiating cells and their neighbours.

Par-1 phosphorylates and thereby excludes Bazooka from the basal-lateral regions of a polarized epithelial cell. We examined the behaviour of Bazooka^{S151A, S1085A}, a mutant form of Bazooka that cannot be phosphorylated by the Par-1 kinase⁴. When the endogenous Bazooka is present, the GFP-tagged Bazooka^{S151A, S1085A} shows a subapical (junctional) distribution similar to the GFP-tagged wild-type form (Fig. 3c and Supplementary Movie 15). However, when we knocked down the endogenous Bazooka using RNAi, Bazooka^{S151A, S1085A} initially shows a broad distribution along the apical–basal axis and eventually coalesces in the lateral regions of all dorsal cells (Fig. 3d and Supplementary Movie 16). A similar localization was observed for wild-type Bazooka-GFP in *par-1* RNAi embryos (Supplementary Movie 17) and in both cases, dorsal fold formation is blocked. These results indicate that serine 151 and 1085 of Bazooka are two main substrates of Par-1 during dorsal fold initiation, whose differential phosphorylation determines the heterogeneous positioning of Bazooka across the dorsal epithelium and is critical for dorsal fold initiation.

When we altered the ratio of Bazooka/Par-1 by a uniform increase in Bazooka levels throughout the epithelium, we saw shifts of junctions that lead to eventual formation of epithelial folds in regions that are outside the sites of anterior and posterior folds and typically near the third and seventh stripes of Runt expression (Fig. 3e, Supplementary Movie 18 and Supplementary Fig. 10). An increase in the Bazooka levels alone in the cells that normally maintain a subapical positioning of junctions can thus be sufficient to drive junctional shift and epithelial folding, presumably by exploiting subtle local heterogeneities in Par-1 levels (Supplementary Fig. 11).

In most epithelia, aPKC phosphorylates Bazooka and becomes segregated to establish the apical domain above the junctions^{7,9,10}. We asked whether aPKC plays a role during junctional repositioning. In embryos that lack aPKC activity, the basal margin of the junctions

shows its characteristic basal shift in the initiating cells, but the apical margin unexpectedly maintains its typical subapical positioning, leading to an abnormally wide junctional domain. In contrast, the width and positioning of the junctions in the neighbouring cells appear normal (Fig. 4a and Supplementary Movie 19). These results indicate that aPKC controls the apical margin to maintain the size of the junctions, but is not required for the basal shift of junctions. These observations also decouple the junctional shift from an increase in the size of the apical membrane. The widening of junctional expanse was also observed in embryos that overexpress Bazooka^{S980A}, a mutant form of Bazooka that cannot be phosphorylated by aPKC (Supplementary Fig. 12)⁹. It seems that the segregation of aPKC from Bazooka establishes the apical domain, enabling junctional disassembly at the apical margin of the junctions.

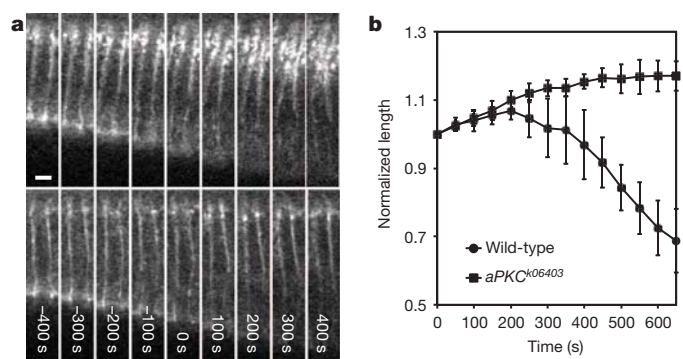


Figure 4 | Loss of aPKC results in an expansion of the junctional domain and a failure to shorten the initiating cells. **a**, Two-photon time-lapse images of E-Cadherin-GFP in an initiating (top) and a neighbouring cell (bottom) in an *aPKC* mutant embryo. Scale bar, 5 μ m. **b**, A time-course analysis of normalized cell length of the initiating cells in the wild-type ($n = 5$) and *aPKC* mutant ($n = 4$) embryos. Error bars indicate s.d.

Although the basal shift of junctions occurs in the initiating cells in the *aPKC* mutant embryo, these cells fail to shorten and the dorsal folds do not form properly (Fig. 4b and Supplementary Movie 19). It seems that as the basal margin of the junctions shifts basally in the initiating cells in response to a decrease in Par-1 levels, their apical margin needs to become disassembled in an *aPKC*-dependent manner so that the subsequent apical cell shape changes could occur.

In this report, we present evidence that dorsal fold initiation requires the establishment of distinct ratios of Bazooka/Par-1 that impose different positions for the adherens junctions in the initiating cells and their neighbours. We propose that the differential positioning of junctions facilitates epithelial folding through two cellular processes (Supplementary Fig. 13). Within the initiating cells, the resultant increase in the non-adherent apical surface after junctional shift may be unstable such that a shrinkage of the apical domain is triggered to restore the balance between cell surface tension and local adhesive forces¹¹. The shortened cells thus produced would then create a localized structural inhomogeneity in the epithelium where buckling would preferentially occur. Second, in the immediate flanking cells, a junctional asymmetry is produced because the basal positioning in the initiating cells on one side and the subapical positioning in the neighbouring cells on the other must be accommodated. Because all junctions in an epithelium are mechanically coupled¹², the asymmetry may cause the lateral surfaces to curve and cells to bend towards the shortened initiating cells. This bending would drive and deepen any buckles or folds initiated in the epithelial sheet.

Directional movement of the cadherin complex along the apical-basal axis has been observed previously in cultured cells *in vitro*¹³ but, to our knowledge, *Drosophila* gastrulation provides the first case where such movement has been described in an intact developing organism. When the shifts occur in stripes as they do on the dorsal side of the *Drosophila* embryo, they seem to initiate infolding of the epithelium. In tissues in which the levels of cortical myosin are low and constant, junctional repositioning regulated by Par-1/Bazooka interactions may play a more prominent role in epithelial folding than does differential activation of cortical contractility. Junctional repositioning may also represent an important mechanism in folding events that do not lead to internalization or delamination, or where the integrity of junctions within the epithelia must be maintained. How junctions are repositioned while maintaining junctional integrity is unclear, but in principle the process could involve remodelling via local endocytic trafficking, or lateral movement of the intact junctions in the membrane¹⁴. Regardless of the mechanism, dorsal fold formation represents an emergent model in which the insights into this alternative mode of epithelial folding could be further analysed.

METHODS SUMMARY

Detailed information about reagents and methods used in this paper, including the *Drosophila* stocks, RNAi, live imaging, immunofluorescence, scanning electron microscopy, image processing, three-dimensional cell boundary reconstruction and image quantification is described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 May 2011; accepted 9 February 2012.

Published online 28 March 2012.

1. Sawyer, J. M. *et al.* Apical constriction: a cell shape change that can drive morphogenesis. *Dev. Biol.* **341**, 5–19 (2010).
2. Sweeton, D., Parks, S., Costa, M. & Wieschaus, E. Gastrulation in *Drosophila*: the formation of the ventral furrow and posterior midgut invaginations. *Development* **112**, 775–789 (1991).
3. Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457**, 495–499 (2009).
4. Benton, R. & St Johnston, D. *Drosophila* PAR-1 and 14-3-3 inhibit Bazooka/PAR-3 to establish complementary cortical domains in polarized cells. *Cell* **115**, 691–704 (2003).
5. Harris, T. J. & Peifer, M. Adherens junction-dependent and -independent steps in the establishment of epithelial cell polarity in *Drosophila*. *J. Cell Biol.* **167**, 135–147 (2004).
6. Harris, T. J. & Peifer, M. The positioning and segregation of apical cues during epithelial polarity establishment in *Drosophila*. *J. Cell Biol.* **170**, 813–823 (2005).
7. Krahn, M. P., Buckers, J., Kastrup, L. & Wodarz, A. Formation of a Bazooka–Stardust complex is essential for plasma membrane polarity in epithelia. *J. Cell Biol.* **190**, 751–760 (2010).
8. McGill, M. A., McKinley, R. F. & Harris, T. J. Independent cadherin–catenin and Bazooka clusters interact to assemble adherens junctions. *J. Cell Biol.* **185**, 787–796 (2009).
9. Morais-de-Sá, E., Mirose, V. & St Johnston, D. *aPKC* phosphorylation of Bazooka defines the apical/lateral border in *Drosophila* epithelial cells. *Cell* **141**, 509–523 (2010).
10. Walther, R. F. & Pichaud, F. Crumbs/DaPKC-dependent apical exclusion of Bazooka promotes photoreceptor polarity remodeling. *Curr. Biol.* **20**, 1065–1074 (2010).
11. Lecuit, T. & Lenne, P. F. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. *Nature Rev. Mol. Cell Biol.* **8**, 633–644 (2007).
12. Martin, A. C., Gelbart, M., Fernandez-Gonzalez, R., Kaschube, M. & Wieschaus, E. F. Integration of contractile forces during tissue invagination. *J. Cell Biol.* **188**, 735–749 (2010).
13. Kametani, Y. & Takeichi, M. Basal-to-apical cadherin flow at cell junctions. *Nature Cell Biol.* **9**, 92–98 (2007).
14. Cavey, M. & Lecuit, T. Molecular bases of cell-cell junctions stability and dynamics. *Cold Spring Harb. Perspect. Biol.* **1**, a002998 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Martin, J. McDonald, D. St Johnston and J. Zallen for providing flies and antibodies; J. Goodhouse and S. Thiberge for assistance in microscopy; G. Deshpande, members of the Wieschaus and Schüpbach laboratories for helpful comments on the manuscript and discussion. This work is supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation to Y.-C.W., a National Institutes of Health/National Institute of General Medical Sciences P50 grant (GM071508) to Z.K. and M.K., and a National Institute of Child Health and Human Development grant (5R37HD15587) to E.F.W. E.F.W. is an investigator of the Howard Hughes Medical Institute.

Author Contributions Y.-C.W. conceived the project, performed all experiments and analysed the data, except those experiments that involve scanning electron microscopy, which were performed by E.F.W.; Z.K. and M.K. developed the software for three-dimensional reconstruction. Y.-C.W. and E.F.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.F.W. (efw@princeton.edu).

METHODS

Drosophila genetics. *Drosophila* stocks used for live imaging were: *Resille-GFP*¹⁵, *myosin-GFP* (*Spaghetti squash-GFP*)¹⁶, *membrane-mCherry*¹² (also known as *P{sqh-mCherry.membrane}*), *E-Cadherin-GFP*¹⁷, *Par-1 protein trap*¹⁸ (also known as *P{PTT-GC}par-1^{CC01981}*), *mat-tub-GFP-Par-1* (ref. 19). *UASp-Bazooka-GFP*²⁰ was driven maternally by one copy (67C) of the *matzTub-Gal4VP16* driver in live imaging experiments and one or two copies (67C; 15) in overexpression experiments. *UASp-Bazooka^{S151A, S1085A}-GFP*⁴ and *UASp-Bazooka^{S980A}-GFP*⁹ were driven by *matzTub-Gal4VP16* (67C; 15). Mutant stocks used were: *runt^{LB5}*, *torso^{XR}/torso^{PM51}*, *torso-like¹/torso-like⁴*, *bicoid^{E1}* *nanos^{BN}* *torso-like³*. Germline clones of *bazooka^{GD21}* and *aPKC^{K06402}* were generated using the FLP-recombinase/dominant female sterile system with the *ovo^{D2} FRT^{19A}* or *FRT^{G13} ovo^{D1}* chromosome (*FRT^{19A}* and *FRT^{G13}* are also known as *P{neoFRT}19A* and *P{FRT(w^{hS})}G13*).

RNAi. Double-stranded RNAs were synthesized using a MEGascript T7 kit (Ambion) for PCR products that contain the T7 promoter sequence (5'-TA ATACGACTCACTATAGGGTACT-3') at each end. The PCR products used in *in vitro* transcription reactions were amplified from 0–4 h embryonic cDNA using the following primer pairs: *bazooka*, 5'-GACGTTTCTTGCTAAGCGG-3', 5'-TTTCGCAGTGTAGGTCCAAA-3'; *bazooka* 5'UTR (for knockdown of endogenous but not transgenic *bazooka*), 5'-AATGCGCGCTGTATGAATCA CAC-3', 5'-ACGACCGCATCATCATCGTCG-3'; *par-1*, 5'-CACGTTCTG CCGTAGCC-3', 5'-GCTTGGGATCGGCTAAATC-3. Double-stranded RNAs were injected into the embryos during the syncytial blastoderm stage, typically 3–4 h before imaging.

Live imaging, immunofluorescence and scanning electron microscopy. Two-photon live embryo imaging was performed on a custom-made system built on an upright Olympus BX51 microscope that is equipped with a Ti:sapphire tunable laser ranged from 720 to 960 nm (Coherent). Single-photon confocal imaging was performed on a Leica SP5 system. Immunofluorescence was performed on heat-methanol fixed embryos²¹. Antibodies used were mouse monoclonal anti-Neurotactin (BP106, Developmental Studies Hybridoma Bank, 1:20), rabbit anti-Runt (1:1,000), rabbit anti-Armadillo (1:200), rabbit anti-Par-1 (ref. 22, 1:500), rabbit anti-PKC ζ C20 (1:1,000, Santa Cruz Biotechnology), and guinea pig anti-Bazooka²³ (1:500) and were visualized by Alexa 488-, 568- and 647-conjugated secondary antibodies (Molecular Probes). Scanning electron microscopy was performed on a Hitachi TM-1000 system as previously described¹². Images were processed, assembled into figures and converted into movies using ImageJ, Adobe Photoshop and Adobe Illustrator.

Three-dimensional image processing and cell boundary reconstruction. The algorithm for three-dimensional reconstruction and analysis was implemented in C++ using the Qt library and OpenGL for the graphical user interface. Libtiff was used for loading image stacks. Image stacks of three-channel immunofluorescence were used. All channels of the 8-bit image volume were initially scaled to a 1:1:1 aspect ratio (voxel size, 0.16 $\mu\text{m} \times 0.16 \mu\text{m} \times 0.16 \mu\text{m}$) and down-sampled by 80% (voxel size, 0.20 $\mu\text{m} \times 0.20 \mu\text{m} \times 0.20 \mu\text{m}$) to reduce image noise. The contrast of Neurotactin immunofluorescence was enhanced by adaptive histogram adjustment using eight equally spaced histograms along the z-dimension and the 10th percentile as the minimum intensity value for intensity adjustment. Two passes of rank filtering were used to fill weak regions of Neurotactin staining, closing holes between cells: a rank filter using the 95th percentile intensity value in a sphere with radius 0.6 μm , followed by a rank filter using the 10th percentile intensity value in a sphere with radius 0.8 μm . Edges were found using a difference of Gaussians (DOG) approximation to a three-dimensional Marr–Hildreth edge detector where the zero-crossing was positioned at low threshold of 4 and high threshold of 30 to generate two binary image volumes²⁴. A “rolling ball” algorithm applied to the high-threshold volume was used to repair holes in the epithelium due to ongoing cellularization. The algorithm was computed efficiently using boundaries in a Euclidian distance transform (EDT)²⁵. Briefly, a boundary at distance of 3 μm was defined using the EDT. A second boundary, also at 3 μm from the first boundary, was used to approximate the result of rolling a sphere on the high-threshold binary image generated from the Marr–Hildreth operator. The repaired boundary was applied to the low-threshold binary image to obtain a binary image where the outer boundary of the epithelium was repaired. This binary image was then thinned by three-dimensional surface thinning²⁶. Connected components in the surface-thinned binary image were found by depth first search. Components of fewer than 100 voxels were removed as noise. Cells were found in this image by hierarchical application of a seeded Watershed algorithm^{27,28}. Seed regions were defined hierarchically by gradually applying an increasing threshold to an EDT of the thinned binary image until a Watershed-segmented region reached a volume of less than 640 μm^3 . Regions smaller than 40 μm^3 were removed as noise. The segmented cell regions were then converted into three-dimensional triangle meshes by the Marching Cubes algorithm²⁹. Lastly, the resulting meshes were

adapted to the intensity of the image by a finite difference approximation to an Active Surface³⁰.

Image quantification. The frequency of myosin bursts (Supplementary Fig. 3b) was measured in confocal time-lapse images of myosin-GFP and membrane-mCherry. These data sets have a z resolution of 1 μm and cover a 10 μm distance from the apical cortex with a temporal resolution ranging between 15 and 22 s per frame. The intense myosin structures were visually identified from each z slice of the image stacks. The total number of myosin bursts was the sum of myosin bursts from all z slices throughout the duration of imaging. The frequency was then calculated by dividing the total number of bursts by the imaging duration and the number of cells in which the bursts were counted. The imaging duration ranges between 436 and 689 s. The numbers of initiating and neighbouring cells that were counted range between 30 to 41 and 51 to 86, respectively.

Time-course analyses of junctional positioning and apical domain size (Supplementary Fig. 4, 6) were performed using two-photon time-lapse images of E-Cadherin-GFP and Resille-GFP. The onset of gastrulation was defined by the onset of anterior cell movement driven by the posterior midgut invagination. For the analysis of junctional positioning, the central initiating cell of the anterior and posterior folds and a third cell that resides in the region between the anterior and posterior fold that shows minimal junctional movement were chosen to represent the initiating and neighbouring cells. The vertical distance between the visually defined centre of the junctional complex and the apex of the cell was measured in Image J to represent the positioning of the junctions. For the analysis of apical domain size, the central initiating cell of the posterior fold and a representative neighbouring cell in the region between the anterior and posterior folds were chosen. The apical domain above the junctions was manually outlined based on the membrane fluorescence of Resille-GFP and measured for its area and perimeter in ImageJ.

Correlation analysis between differential junctional displacement and lateral membrane curvature (Supplementary Fig. 5) was performed using two-photon time-lapse images of E-Cadherin-GFP. Cells that are in the immediate flanking regions of the initiating cells and show a marked asymmetry of junctional positioning on the opposite sides were chosen for these measurements. The differential junctional displacement, which defines the extent of junctional asymmetry, was calculated by subtracting the length of the apical domain on the distal side (y) from that on the proximal side (x). The lateral membrane curvature was defined as the ratio between the height (h) and the chord (C) of the arc of lateral membrane on the distal side of the cell.

Three-dimensional cell shape measurements (Fig. 1j–l) were made in image stacks of late cellularization embryos that have been stained for Bazooka and Neurotactin and processed for three-dimensional reconstruction as described above. The position of Bazooka was defined by the ‘Bazooka junctional triangles’. Briefly, an average intensity of Bazooka was first assigned for the voxels that intersect with a three-dimensional triangle mesh in the reconstructed cell boundary. The Bazooka junctional triangles were then selected based on an intensity threshold of the 99th percentile of the Bazooka intensity histogram. For each of the centroids of the Bazooka junctional triangles, a three-dimensional principal component analysis (PCA) was performed to determine the Bazooka mean position (a point on a plane) and the eigenvector corresponding to smallest eigenvalue (plane normal). These were then used to define the Bazooka junctional plane that subdivides the cell into the apical and basal domains. The geometric measurements were performed as follows: three-dimensional PCA was applied to all of the vertices of the triangle mesh and the long direction of the cell was defined using the eigenvector corresponding to the largest eigenvalue. The apical domain length was measured by first creating vectors between the cell centroid and each mesh vertex on the apical side of the Bazooka junctional plane. These vectors were then projected onto the long direction vector of the cell. The length of the longest projected vector was used as the apical domain length. The basal domain length was measured similarly, using triangle mesh vertices on the basal domain of the cell. The total cell length was computed by a sum of the apical and basal length. The apical volume was computed by voxelizing the three-dimensional triangle meshes, and summing the volumes of voxels apical to the Bazooka junctional plane. Similarly, the apical surface area was computed by summing the areas of the mesh triangles for which the triangle centroid falls on the apical side of the Bazooka junctional plane. The initiating cells were selected on the basis of their location and junctional positioning. Approximately 500 dorsal cells in the region between the first and the seventh stripe of Runt were ranked by the apical domain length and the top 150 cells were selected for further analysis. A second selection was performed to isolate those that are in close proximity to the second and fifth stripes of Runt. Of these cells, those whose apical domain length was above the average were used for analysis. For the early stage embryos that showed no junctional shift, only the location-based selection was made. Cells that reside in the region between the anterior and posterior folds with junctional positioning that was below average were used as the neighbouring cells.

Time-course analysis of Par-1 dynamics (Fig. 2c) was performed using two-photon time-lapse images of Par-1-GFP. The average fluorescent intensity of Par-1-GFP in manually selected areas consisting of two anterior or four posterior initiating cells was measured and normalized by that in areas consisting of four neighbouring cells that reside in the regions between the anterior and posterior folds. The onset of gastrulation was defined by the onset of anterior cell movement driven by posterior midgut invagination.

Bazooka and Par-1 immunofluorescence (Fig. 2d, e and Supplementary Fig. 11) was quantified in image stacks of fixed embryos that were triply labelled for Bazooka, Par-1 and Neurotactin and processed for three-dimensional reconstruction as described above to define the Bazooka junctional triangles. Junctional intensity of Bazooka within a cell was measured within and normalized by the Bazooka junctional volume that was defined by the voxelization of Bazooka junctional triangles. The basal-lateral intensity of Par-1 in each cell was measured within and normalized by the volume within a two-voxel distance from the cell boundary basal to the Bazooka junctional plane. For Fig. 2d, e, the anterior and posterior fold-initiating cells were selected on the basis of location and above-average Bazooka positioning, whereas the neighbouring cells were selected from the cells that reside in the region between the anterior and posterior folds with below-average Bazooka positioning. For Supplementary Fig. 11, the wild-type and Bazooka overexpression embryos were fixed, stained and imaged in parallel under identical conditions. Cell selection was performed as in Fig. 2d, e.

Time-course of cell shortening in the initiating cells (Fig. 4b) was analysed by measuring the total cell length of five and four initiating cells each from wild-type and *aPKC* mutant embryos in ImageJ. The measurements were normalized by cell length at time zero, which was defined by the onset of anterior cell movement driven by the posterior midgut invagination.

15. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
16. Royou, A., Sullivan, W. & Karess, R. Cortical recruitment of nonmuscle myosin II in early syncytial *Drosophila* embryos: its role in nuclear axial expansion and its regulation by Cdc2 activity. *J. Cell Biol.* **158**, 127–137 (2002).
17. Oda, H. & Tsukita, S. Real-time imaging of cell-cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114**, 493–501 (2001).
18. Lighthouse, D. V., Buszczak, M. & Spradling, A. C. New components of the *Drosophila* fusome suggest it plays novel roles in signaling and transport. *Dev. Biol.* **317**, 59–71 (2008).
19. Shulman, J. M., Benton, R. & St Johnston, D. The *Drosophila* homolog of *C. elegans* PAR-1 organizes the oocyte cytoskeleton and directs *oskar* mRNA localization to the posterior pole. *Cell* **101**, 377–388 (2000).
20. Benton, R. & St Johnston, D. A conserved oligomerization domain in *Drosophila* Bazooka/Par-3 is important for apical localization and epithelial polarity. *Curr. Biol.* **13**, 1330–1334 (2003).
21. Müller, H. A. & Wieschaus, E. *armadillo*, *bazooka*, and *stardust* are critical for early stages in formation of the zonula adherens and maintenance of the polarized blastoderm epithelium in *Drosophila*. *J. Cell Biol.* **134**, 149–163 (1996).
22. McDonald, J. A., Khodyakova, A., Aranjuez, G., Dudley, C. & Montell, D. J. PAR-1 kinase regulates epithelial detachment and directional protrusion of migrating border cells. *Curr. Biol.* **18**, 1659–1667 (2008).
23. Simões, M. et al. Rho-kinase directs Bazooka/Par-3 planar polarity during *Drosophila* axis elongation. *Dev. Cell* **19**, 377–388 (2010).
24. Marr, D. & Hildreth, E. Theory of edge detection. *Proc. R. Soc. Lond. B Biol. Sci.* **207**, 187–217 (1980).
25. Maurer, C. R. Jr, Qi, R. & Raghavan, V. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 265–270 (2003).
26. Palágyi, K. et al. A sequential 3D thinning algorithm and its medical applications information processing in medical imaging. *Lect. Notes Comput. Sci.* **2082**, 409–415 (2001).
27. Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598 (1991).
28. Beucher, S. & Meyer, F. In *Mathematical Morphology in Image Processing* Ch. 12 (ed. Dougherty, E. R.) 433–481 (Marcel Dekker, 1993).
29. Lorensen, W. E. & Cline, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH '87: Proceedings of the 14th annual conference on computer graphics and interactive techniques* **21**, 163–169 (1987).
30. Cohen, L. D. & Cohen, I. Finite element methods for active contour models and balloons for 2D and 3D images. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1131–1147 (1991).

Hsp72 preserves muscle function and slows progression of severe muscular dystrophy

Stefan M. Gehrig¹, Chris van der Poel^{1†}, Timothy A. Sayer¹, Jonathan D. Schertzer^{1†}, Darren C. Henstridge², Jarrod E. Church^{1†}, Severine Lamon³, Aaron P. Russell³, Kay E. Davies⁴, Mark A. Febbraio² & Gordon S. Lynch¹

Duchenne muscular dystrophy (DMD) is a severe and progressive muscle wasting disorder caused by mutations in the dystrophin gene that result in the absence of the membrane-stabilizing protein dystrophin^{1–3}. Dystrophin-deficient muscle fibres are fragile and susceptible to an influx of Ca²⁺, which activates inflammatory and muscle degenerative pathways^{4–6}. At present there is no cure for DMD, and existing therapies are ineffective. Here we show that increasing the expression of intramuscular heat shock protein 72 (Hsp72) preserves muscle strength and ameliorates the dystrophic pathology in two mouse models of muscular dystrophy. Treatment with BGP-15 (a pharmacological inducer of Hsp72 currently in clinical trials for diabetes) improved muscle architecture, strength and contractile function in severely affected diaphragm muscles in *mdx* dystrophic mice. In *dko* mice, a phenocopy of DMD that results in severe spinal curvature (kyphosis), muscle weakness and premature death^{7,8}, BGP-15 decreased kyphosis, improved the dystrophic pathophysiology in limb and diaphragm muscles and extended lifespan. We found that the sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase (SERCA, the main protein responsible for the removal of intracellular Ca²⁺) is dysfunctional in severely affected muscles of *mdx* and *dko* mice, and that Hsp72 interacts with SERCA to preserve its function under conditions of stress, ultimately contributing to the decreased muscle degeneration seen with Hsp72 upregulation. Treatment with BGP-15 similarly increased SERCA activity in dystrophic skeletal muscles. Our results provide evidence that increasing the expression of Hsp72 in muscle (through the administration of BGP-15) has significant therapeutic potential for DMD and related conditions, either as a self-contained therapy or as an adjuvant with other potential treatments, including gene, cell and pharmacological therapies.

DMD is the most severe form of muscular dystrophy; it affects about 1 in 3,500 live male births¹. Intracellular Ca²⁺ regulation is compromised in dystrophic muscle fibres, which triggers chronic inflammation, repeated cycles of degeneration with progressively ineffective regeneration, and infiltration of fibrotic and other non-contractile material⁴. Mechanisms for the influx of Ca²⁺ into dystrophic muscle fibres include membrane tears^{4–6}, stretch-activated channels⁹, Ca²⁺ leak channels¹⁰ and leaky Ca²⁺ release channels¹¹, and it has been speculated that the function of SERCA, the main protein responsible for Ca²⁺ reuptake into the sarcoplasmic reticulum (SR), is compromised^{12,13}. Increasing SERCA pump expression within dystrophic muscles in transgenic mice or through viral-mediated delivery improves Ca²⁺ handling and suppresses the pathological cascade of events^{14,15}. The role of inflammation in the dystrophic pathology is well known, particularly that of the pro-inflammatory cytokine tumour necrosis factor- α (TNF- α)¹⁶. TNF- α activates the nuclear factor- κ B (NF- κ B) and c-Jun N-terminal kinase (JNK) signalling pathways^{17–19}. Hsp72 is a molecular chaperone protein that inhibits inflammatory mediators including p-JNK, TNF- α and the

NF- κ B pathway^{20–22}, and binds and preserves SERCA function under conditions of cellular stress²³. Although some studies have shown Hsp72 to be elevated in patients with DMD²⁴, there is little consensus, because expression data for young patients are variable and sourcing age-matched controls is problematic (Supplementary Fig. 1a). Nevertheless, the endogenous heat shock protein response in DMD is insufficient to be protective. Here we tested the hypothesis that increasing the levels of Hsp72 protects dystrophic muscles from functional deterioration. We bred dystrophin-null *mdx* dystrophic mice with mice showing a muscle-specific transgenic (TG) overexpression of Hsp72, producing *mdx*^{TG(+)} mice and *mdx* littermate controls (Fig. 1a; see Supplementary Fig. 1b for quantification). At about 25 weeks of age, serum levels of creatine kinase (CK), a classic indicator of muscle breakdown, were decreased in *mdx*^{TG(+)} mice compared with littermate control mice lacking the transgene (*mdx*^{TG(–)}; Fig. 1b). Because most patients with DMD show severe weakness⁴ and/or muscle fatigue²⁵, we assessed whole-body strength and endurance in dystrophic mice by using a hang test to measure latency-to-fall, which was significantly improved in *mdx*^{TG(+)} mice (24 ± 3 s versus 64 ± 14 s; $P = 0.002$, $n \geq 20$ mice). Respiratory failure is the cause of death in up to 90% of patients with DMD²⁶, and because diaphragm function is an accurate predictor of respiratory insufficiency we investigated the effect of Hsp72 overexpression on the pathophysiology of the diaphragm in dystrophin-deficient mice. The progressive degeneration of the diaphragm in *mdx* mice closely mimics that in DMD²⁷. Gross histological analyses revealed that the diaphragm pathology in *mdx*^{TG(+)} mice was ameliorated compared with age-matched littermate control *mdx*^{TG(–)} mice (Supplementary Fig. 1c), an observation supported by the minimal Feret's diameter variance coefficient, which provides a sensitive measure of fibre heterogeneity and the dystrophic pathology²⁸. We found a lower Feret's diameter variance coefficient in *mdx*^{TG(+)} mice than in *mdx*^{TG(–)} mice, indicative of an improved phenotype (405 ± 5 versus 358 ± 5 ; $P < 0.001$, $n = 5$). Damaged myofibres can be revealed by the infiltration of Evans blue dye (EBD) entering the myoplasm through tears in the sarcolemma^{4–6}. EBD infiltration was decreased in cross-sections of the diaphragm from *mdx*^{TG(+)} mice in comparison with that from *mdx*^{TG(–)} mice (Fig. 1c and Supplementary Fig. 1d), indicative of decreased necrosis and hastened overall repair of damaged fibres rather than improved structural integrity. To support this contention, we performed well-described *in situ* and *in vitro* contraction-induced injury protocols on tibialis anterior (TA) and diaphragm muscles, respectively (see Methods). No differences in contraction-mediated damage were evident between muscles of *mdx*^{TG(+)} and *mdx*^{TG(–)} mice (Supplementary Fig. 1e), indicating that structural integrity was unaltered. Expression of the dystrophin homologue, utrophin, a protein known to compensate for the loss of dystrophin, was also unchanged (Supplementary Fig. 1f). We assessed collagen infiltration in sections of

¹Basic and Clinical Myology Laboratory, Department of Physiology, University of Melbourne, Victoria, 3010, Australia. ²Cellular and Molecular Metabolism Laboratory, Baker-IDI Heart and Diabetes Institute, PO Box 6492, St Kilda Road Central, Victoria, 8008, Australia. ³Centre for Physical Activity and Nutrition Research, School of Exercise and Nutrition Sciences, Deakin University, Burwood, Victoria, 3125, Australia. ⁴MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. [†]Present addresses: Department of Human Biosciences, Faculty of Health Sciences, La Trobe University, Bundoora, 3086, Victoria, Australia (C.v.d.P. and J.E.C.); Department of Biochemistry and Biomedical Sciences and Department of Pediatrics, McMaster University, Hamilton, Ontario, L8S 4L8, Canada (J.D.S.).

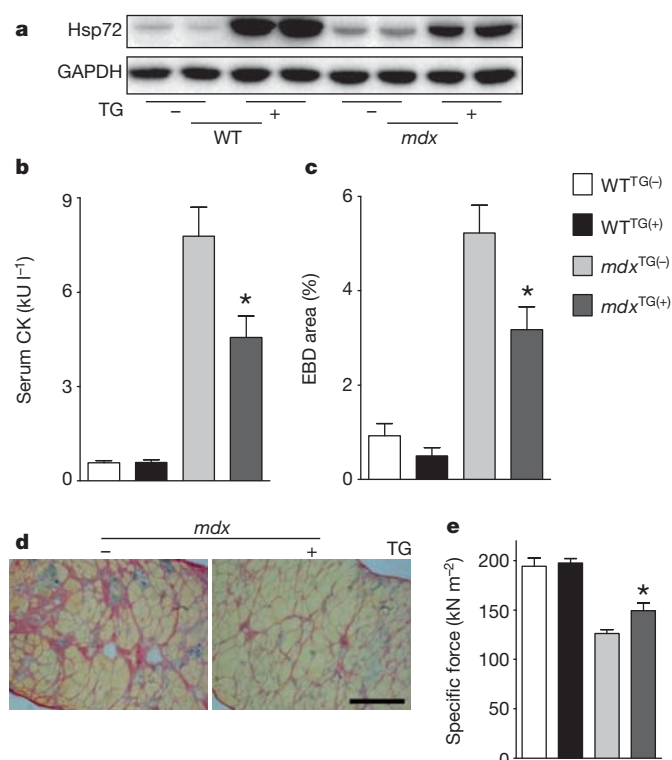


Figure 1 | Transgenic Hsp72 overexpression increases muscle strength, decreases muscle breakdown and improves diaphragm muscle histological parameters in *mdx* mice. **a**, Representative western blot detection of Hsp72 in diaphragm muscle homogenates from non-dystrophic (WT) and dystrophic (*mdx*) Hsp72 transgene-negative (TG-) and transgene-positive (TG+) mice. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. **b**, Whole-body muscle breakdown, measured by serum creatine kinase (CK) levels. Asterisk, $P < 0.05$, $n \geq 6$. **c**, Quantified mean data for EBD-positive area in diaphragm muscle sections. Asterisk, $P < 0.05$, $n \geq 4$. **d**, Representative images of collagen infiltration in diaphragm sections (revealed with Van Gieson's stain) from *mdx*^{TG(-)} and *mdx*^{TG(+)} mice. **e**, Specific (normalized) force of diaphragm muscle strips measured *in vitro*. Asterisk, $P < 0.05$, $n \geq 5$. All data are from 25-week-old mice. Scale bar, 200 μ m. Data are shown as means \pm s.e.m.

diaphragm and found that *mdx*^{TG(+)} mice had less collagen infiltration than *mdx*^{TG(-)} mice at both 25 weeks of age (Fig. 1d and Supplementary Fig. 1g) and 80 weeks of age (data not shown). Normalized force production was significantly higher in diaphragm muscle strips from *mdx*^{TG(+)} mice than in those from *mdx*^{TG(-)} mice (Fig. 1e).

We have shown that Hsp72 can block inflammation in mice *in vivo*²², and because inflammation promotes muscle degeneration in mouse models of DMD¹⁷ we examined the effect of Hsp72 on expression of p-JNK, p-IKK (a key mediator of NF- κ B activation) and TNF- α . Western blot and polymerase chain reaction (PCR) analyses revealed no difference in these inflammatory markers in diaphragm muscles from *mdx*^{TG(-)} and *mdx*^{TG(+)} mice (Supplementary Fig. 2, and data not shown). There was significantly decreased messenger RNA expression of macrophage markers CD68 and F4/80 and TNF- α in TA muscles of *mdx*^{TG(+)} mice (Supplementary Fig. 3), but these did not translate to functional improvements (data not shown). We next examined SERCA activity in diaphragm muscles from wild-type (WT) and *mdx* mice and showed a progressive age-related decline in maximal SERCA activity in *mdx* mice (Fig. 2a and Supplementary Fig. 4a), despite an age-related increase in SERCA protein expression (Supplementary Fig. 4b). This functional decline is attributed, in part, to post-translational modifications of the SERCA protein, especially nitrosylation, which decreases maximal SERCA activity as a result of changes in Ca²⁺-binding and ATP-binding domains²³. Alterations in the Ca²⁺-binding domain changes SERCA

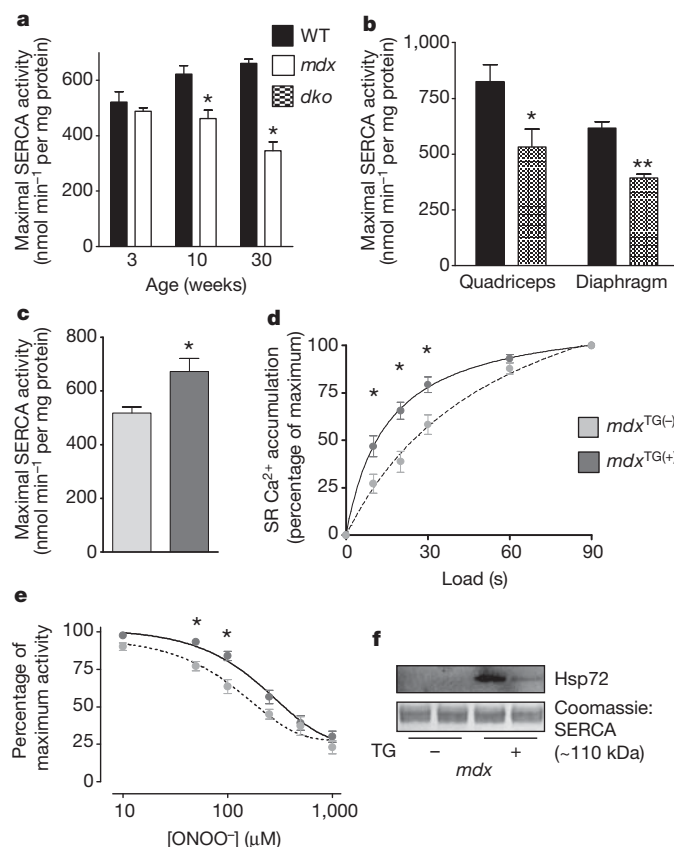


Figure 2 | Maximal SERCA activity is decreased in mouse models of dystrophy; Hsp72 binding improves SERCA function. **a**, Maximal SERCA activity in diaphragm muscle homogenates from WT and *mdx* mice at 3, 10 and 30 weeks of age. Asterisk, $P < 0.001$, $n \geq 4$. **b**, Maximal SERCA activity of muscle homogenates from quadriceps and diaphragm from 10-week-old WT and *mdx* mice. Asterisk, $P = 0.036$. Two asterisks, $P < 0.001$, $n \geq 4$. **c**, Maximal SERCA activity in muscle homogenates from *mdx*^{TG(-)} (pale grey bars) and *mdx*^{TG(+)} (dark grey bars) mice. Asterisk, $P < 0.008$, $n \geq 9$. **d**, Ca²⁺ accumulation curves (an indirect measure of SERCA activity) for SR in single fibres from extensor digitorum longus muscles of *mdx*^{TG(-)} and *mdx*^{TG(+)} mice. Asterisk, $P < 0.01$, $n \geq 15$. **e**, Maximal SERCA activity after incubation for 5 min at various concentrations of peroxynitrite (ONOO⁻), measured in enriched SR vesicles isolated from muscles of *mdx*^{TG(-)} and *mdx*^{TG(+)} mice. SERCA activity was normalized to maximal activity in the absence of ONOO⁻. Asterisk, $P < 0.05$, $n \geq 3$. **f**, Representative Western blots of enriched SR vesicles from *mdx*^{TG(-)} and *mdx*^{TG(+)} ($n = 4$) mice, showing Hsp72 protein expression and Coomassie blue stain (showing SERCA isoforms). Data are shown as means \pm s.e.m.

enzyme kinetics, decreasing Ca²⁺ sensitivity and increasing [Ca²⁺]₅₀ (the [Ca²⁺] required to achieve half-maximal enzyme activity), an effect we observed in diaphragm muscles from 30-week-old *mdx* mice (Supplementary Fig. 4c). Similar deficits in SERCA activity were evident in both limb and diaphragm muscles of severely affected *dco* mice (Fig. 2b), indicating that abnormal SERCA function may contribute to the disruptions in Ca²⁺ regulation characteristic of dystrophic muscles. Indeed, recent evidence suggests that closer regulation of Ca²⁺ homeostasis through enhanced SERCA expression or activity significantly suppresses the degeneration of dystrophic muscle^{14,15}. Because Hsp72 binds SERCA and prevents functional inactivation under conditions of cellular stress²³, we next tested the hypothesis that overexpression of Hsp72 in dystrophic muscles would improve SERCA function. We examined maximal SERCA activity in homogenates of quadriceps muscles, and found an increase in activity in *mdx*^{TG(+)} compared with *mdx*^{TG(-)} mice (Fig. 2c). To support this finding we also examined Ca²⁺ accumulation in SR (an indirect measure of SERCA activity) in single muscle fibres dissected from fast-twitch

extensor digitorum longus, and predominantly slow-twitch soleus muscles, and found a similar increase in maximal SERCA activity in individual muscle fibres in $mdx^{TG(+)}$ compared with $mdx^{TG(-)}$ mice (as for whole-muscle homogenates), with no changes in mRNA for SERCA or in protein expression (Fig. 2d, Supplementary Fig. 5 and Supplementary Table 1). We then examined whether Hsp72 overexpression could protect SERCA function under conditions of stress, like that induced by reactive oxygen–nitrogen species such as peroxynitrite ($ONOO^-$). We developed a SERCA activity assay in which enriched SR vesicles isolated from $mdx^{TG(-)}$ and $mdx^{TG(+)}$ mice were incubated with increasing concentrations of $ONOO^-$; the subsequent suppression in activity was normalized to that in the absence of $ONOO^-$. We found that SERCA activity was greater in the presence of various $ONOO^-$ concentrations in enriched SR vesicles from $mdx^{TG(+)}$ compared with $mdx^{TG(-)}$ mice (Fig. 2e), with western blots revealing Hsp72 protein levels were highly elevated in enriched SR vesicles from $mdx^{TG(+)}$ mice (Fig. 2f; see Supplementary Fig. 6 for full blots), indicating that Hsp72 was bound within the SR to mediate this protective effect. These data indicate that Hsp72 overexpression in dystrophic muscles can protect SERCA from inactivating modifications and is a likely mechanism of protection in $mdx^{TG(+)}$ mice.

Given the significant phenotypic improvements in $mdx^{TG(+)}$ mice, especially in the diaphragm, we examined whether similar effects could be achieved through the pharmacological or heat-therapy induction of Hsp72. BGP-15 is a pharmacological inducer of Hsp72 that can protect against obesity-induced insulin resistance²² and is in Food and Drug Administration (FDA)-approved Phase II clinical trials for diabetes²⁹. Dystrophic mdx mice were treated from 4 to 9 weeks of age (5 weeks) or from 4 to 16 weeks of age (12 weeks) with BGP-15 (15 mg kg⁻¹ per day, oral gavage). WT mice were used for comparisons with mdx mice. Hsp72 protein expression was elevated in diaphragm muscle homogenates from BGP-15-treated mdx mice (and heat-therapy-treated mice) compared with untreated mdx mice (Fig. 3a; see Supplementary Fig. 7a for quantification). CK levels were decreased in BGP-15-treated mdx mice compared with untreated mdx mice (Fig. 3b). EBD infiltration was also reduced with long-term BGP-15 treatment (Fig. 3c). Strength and endurance was evaluated with the inverted hang test; WT mice were stronger than mdx mice, and BGP-15-treated mdx mice showed an increased latency-to-fall compared with control (Supplementary Fig. 7). An increase in fibrosis, as seen in DMD⁴, was observed in the mdx diaphragm (compared with WT), and BGP-15 treatment significantly decreased fibrosis (Fig. 3d; see Supplementary Fig. 7 for quantification). Treatment with BGP-15 attenuated the functional deterioration of the diaphragm muscle significantly (Fig. 3e). Maximal SERCA activity in diaphragm homogenates was increased in mdx mice after long-term treatment with BGP-15, indicating a mechanism consistent with that of transgenic Hsp72 overexpression (Fig. 3f). Because elevated core temperature is a potent inducer of heat shock proteins^{22,30}, we also extensively tested this method of heat shock protein induction (see Methods). Similar beneficial effects to those observed with transgenic Hsp72 overexpression and BGP-15 treatment were seen in mdx mice exposed to repeated heat therapy (see Supplementary Fig. 8).

We then investigated whether treatment with BGP-15 was protective in severely affected dystrophic dko mice, the most phenotypically accurate murine model of DMD^{7,8} (see Methods). The dko mice were treated with BGP-15 from 3–4 weeks until 10 weeks of age. For the survival study, dko mice were treated from 3–4 weeks until death or humane killing (as described in Methods). Photographs of control and BGP-15-treated dko mice were taken at 10 weeks, immediately before killing, and after evisceration and staining of the skeleton with alizarin red (Fig. 4a). Data from WT mice (as in Fig. 3) were also used for comparisons with dko mice. Because boys with DMD have significant paraspinal muscle weakness and in many cases severe kyphosis (spinal curvature)^{4,8}, this was quantified in conscious mice (from the 10-week endpoint cohort) by a blinded investigator using a 1–5 index of

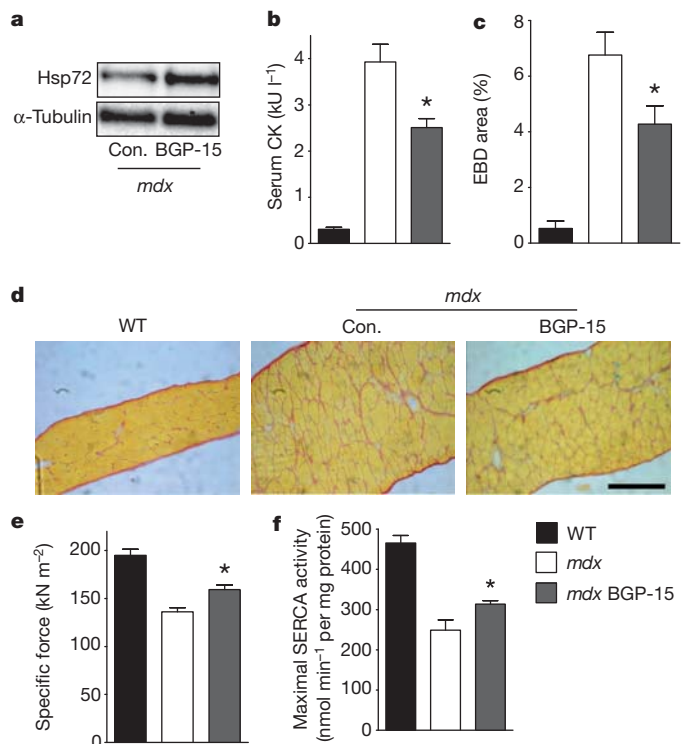


Figure 3 | Pharmacological induction of Hsp72 ameliorates muscular dystrophy in mdx mice. **a**, Representative western blot detection of Hsp72 in diaphragm muscle homogenates of control (Con.) and BGP-15-treated mdx mice. **b**, Whole-body muscle breakdown, measured by serum CK levels in WT (C57BL/10), mdx (control) and BGP-15-treated mdx mice. Asterisk, $P < 0.01$, $n \geq 8$. **c**, Quantified mean data for EBD infiltration in mdx mice treated for 12 weeks. Asterisk, $P < 0.05$, $n \geq 8$. **d**, Representative images of collagen infiltration in diaphragm muscle sections. **e**, Specific force production in diaphragm muscle strips measured *in vitro*. Asterisk, $P < 0.01$, $n \geq 9$. **f**, Maximal SERCA activity in diaphragm muscle homogenates from mdx mice treated for 12 weeks. Asterisk, $P < 0.05$, $n \geq 6$. Scale bar, 200 μ m. WT data are used as a reference control. All treated mice received BGP-15 for 5 weeks unless stated otherwise. Data are shown as means \pm s.e.m.

kyphosis, 1 indicating no spinal deformity on palpation and 5 being the most severe. Treatment with BGP-15 decreased kyphosis markedly compared with vehicle-treated controls (Fig. 4b). Serum CK levels were significantly lower in dko mice after treatment with BGP-15 (Fig. 4c), indicating decreased whole-body muscle breakdown. Collagen infiltration was decreased in the diaphragm of dko mice after treatment with BGP-15 (Fig. 4d) and the force-producing capacity of diaphragm muscle strips and intact TA muscles (measured *in situ*) was increased significantly in BGP-15-treated dko mice, with maximum force restored to WT levels in the TA muscle (Fig. 4e, f and Supplementary Table 2). No differences in body mass or in calcification or central nucleation within diaphragm muscles were evident in dko mice after treatment with BGP-15 (data not shown). However, the most important outcome was that lifelong treatment of dko mice with BGP-15 significantly extended survival (Fig. 4g, h; $P < 0.05$; 27% increase in median lifespan). This finding has clinical relevance for DMD.

Our findings reveal that transgenic Hsp72 overexpression improves several pathological indices in mdx dystrophic mice, at least in part by preserving or improving SERCA function. Furthermore, treatment of dystrophic mice with BGP-15, a known pharmacological co-inducer of Hsp72, ameliorated the dystrophic pathology and extended the lifespan in dko mice. Taken together, these results indicate that induction of Hsp72 in muscular dystrophy is an important and novel therapeutic approach that can improve the dystrophic pathology and attenuate the disease progression. Although an ultimate cure for DMD is likely to be derived from gene or cell therapies, considerable

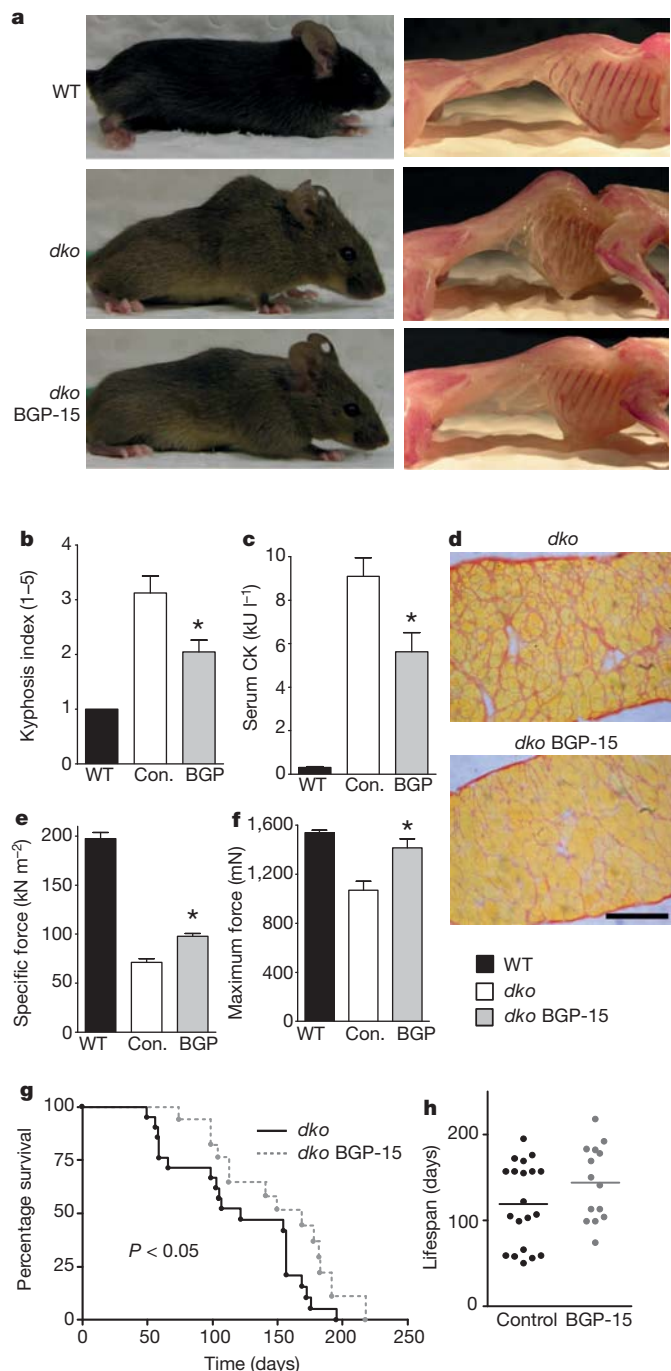


Figure 4 | Treatment with BGP-15 decreases kyphosis (spinal curvature), improves muscle function and prolongs lifespan in severely dystrophic *dko* mice. **a**, Left: representative photographs of WT (C57BL/10), *dko* and BGP-15-treated *dko* mice. Right: representative eviscerated skeletal preparations showing bone structure (pink), and highlighting spinal curvature in *dko* mice. **b**, Spinal curvature was quantified (1–5) in WT, *dko* and BGP-15-treated *dko* mice. Asterisk, $P < 0.05$, $n \geq 10$. **c**, Whole-body muscle breakdown, measured by serum CK levels. Asterisk, $P < 0.05$, $n \geq 8$. **d**, Representative images of collagen infiltration in diaphragm sections. **e**, Specific force of diaphragm muscle strips measured *in vitro*. Asterisk, $P < 0.01$, $n \geq 9$. **f**, Maximal force production in TA measured *in situ*. Asterisk, $P < 0.01$, $n \geq 9$. **g**, Survival curve of untreated (control) and BGP-15-treated *dko* mice. $P < 0.05$, $n \geq 14$. **h**, Scatter-plot of *dko* lifespan with a line showing median survival. Scale bar, 200 μ m. WT data are used as a reference control. Data are shown as means \pm s.e.m.

obstacles need to be overcome before these approaches can be considered safe and effective. Until these concerns are obviated, alternative (and potentially synergistic) therapies, such as pharmacological

induction of Hsp72, could delay the disease progression to allow many patients to benefit from perfected treatments.

In a recent clinical trial for patients with insulin resistance, treatment with BGP-15 (200 mg and 400 mg once a day for 28 days) significantly increased sensitivity to insulin. The dose schedule of BGP-15 was well tolerated, with both doses being safe: there were no clinically significant changes in physical status or in laboratory or electrocardiogram parameters²⁹. Given that BGP-15 is currently used in clinical trials for other pathologies, our findings identify it as a tangible and realistic treatment method for patients with DMD in the near future.

METHODS SUMMARY

Animals. All experiments were approved by the Animal Ethics Committee of The University of Melbourne and conducted in accordance with the Australian code of practice for the care and use of animals for scientific purposes as stipulated by the National Health and Medical Research Council (NHMRC, Australia). Male mice were used for all experiments. Wild-type (C57BL/10) and dystrophic *mdx* mice were sourced from the Animal Resources Centre (Canning Vale, Western Australia).

Muscle functional analysis. Mice were anaesthetized with sodium pentobarbitone (Nembutal) such that they were unresponsive to tactile stimuli. Contractile properties of diaphragm muscle strips were assessed *in vitro*.

Morphological analysis. Muscles were trimmed of tendons and adhering non-muscle tissue, mounted in embedding medium, frozen in liquid-nitrogen-cooled isopentane, and stored at -80°C . Transverse muscle sections were cryosectioned from the mid-belly of each muscle. Muscle collagen content was assessed from Van Gieson's stained cross-sections that were quantified.

SERCA activity assay. Ca^{2+} -dependent SERCA activity was assessed in isolated enriched SR vesicles and whole-muscle homogenates. For whole-muscle homogenates, muscles were surgically excised from anaesthetized mice and stored at -80°C for subsequent analyses. For enriched SR vesicles, mixed hindlimb muscles (quadriceps, gastrocnemius, extensor digitorum longus, soleus and plantaris) and diaphragm muscles were homogenized and subjected to sucrose gradient differential centrifugation using a Thermo Scientific Sorvall WX100 ultracentrifuge with a T-890 fixed-angle rotor. During the entire homogenization and SR vesicle isolation procedures, samples were immersed in ice to avoid temperature-dependent decreases in SERCA activity.

Skeletal preparation. To reveal skeletal architecture, mice (after death) were skinned, eviscerated and placed in a KOH solution (1.5% w/v), for 5 days. KOH solution was replaced and a small amount of alizarin red was added to stain calcium deposits; the preparation was left for a further 5 days.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 January; accepted 21 February 2012.

Published online 4 April 2012.

- Emery, A. E. The muscular dystrophies. *Lancet* **359**, 687–695 (2002).
- Ervasti, J. M. & Campbell, K. P. Membrane organization of the dystrophin-glycoprotein complex. *Cell* **66**, 1121–1131 (1991).
- Koenig, M., Monaco, A. P. & Kunkel, L. M. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53**, 219–228 (1988).
- Blake, D. J., Weir, A., Newey, S. E. & Davies, K. E. Function and genetics of dystrophin and dystrophin-related proteins in muscle. *Physiol. Rev.* **82**, 291–329 (2002).
- Straub, V., Rafael, J. A., Chamberlain, J. S. & Campbell, K. P. Animal models for muscular dystrophy show different patterns of sarcolemmal disruption. *J. Cell Biol.* **139**, 375–385 (1997).
- Turner, P. R., Westwood, T., Regen, C. M. & Steinhardt, R. A. Increased protein degradation results from elevated free calcium levels found in muscle from *mdx* mice. *Nature* **335**, 735–738 (1988).
- Grady, R. M. *et al.* Skeletal and cardiac myopathies in mice lacking utrophin and dystrophin: a model for Duchenne muscular dystrophy. *Cell* **90**, 729–738 (1997).
- Deconinck, A. E. *et al.* Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy. *Cell* **90**, 717–727 (1997).
- Gervasio, O. L., Whitehead, N. P., Yeung, E. W., Phillips, W. D. & Allen, D. G. TRPC1 binds to caveolin-3 and is regulated by Src kinase—role in Duchenne muscular dystrophy. *J. Cell Sci.* **121**, 2246–2255 (2008).
- Fong, P. Y., Turner, P. R., Denetclaw, W. F. & Steinhardt, R. A. Increased activity of calcium leak channels in myotubes of Duchenne human and *mdx* mouse origin. *Science* **250**, 673–676 (1990).
- Bellinger, A. M. *et al.* Hypernitrosylated ryanodine receptor calcium release channels are leaky in dystrophic muscle. *Nature Med.* **15**, 325–330 (2009).
- Nicolas-Metral, V., Raddatz, E., Kucera, P. & Ruegg, U. T. *Mdx* myotubes have normal excitability but show reduced contraction-relaxation dynamics. *J. Muscle Res. Cell Motil.* **22**, 69–75 (2001).

13. Tutdibi, O., Brinkmeier, H., Rudel, R. & Fohr, K. J. Increased calcium entry into dystrophin-deficient muscle fibres of MDX and ADR-MDX mice is reduced by ion channel blockers. *J. Physiol. (Lond.)* **515**, 859–868 (1999).
14. Goonasekera, S. A. *et al.* Mitigation of muscular dystrophy in mice by SERCA overexpression in skeletal muscle. *J. Clin. Invest.* **121**, 1044–1052 (2011).
15. Morine, K. J., Sleeper, M. M., Barton, E. R. & Sweeney, H. L. Overexpression of SERCA1a in the *mdx* diaphragm reduces susceptibility to contraction-induced damage. *Hum. Gene Ther.* **12**, 1735–1739 (2010).
16. Porter, J. D. *et al.* A chronic inflammatory response dominates the skeletal muscle molecular signature in dystrophin-deficient *mdx* mice. *Hum. Mol. Genet.* **11**, 263–272 (2002).
17. Acharyya, S. *et al.* Interplay of IKK/NF- κ B signaling in macrophages and myofibres promotes muscle degeneration in Duchenne muscular dystrophy. *J. Clin. Invest.* **117**, 889–901 (2007).
18. Kolodziejczyk, S. M. *et al.* Activation of JNK1 contributes to dystrophic muscle pathogenesis. *Curr. Biol.* **11**, 1278–1282 (2001).
19. Monici, M. C., Aguenouz, M., Mazzeo, A., Messina, C. & Vita, G. Activation of nuclear factor- κ B in inflammatory myopathies and Duchenne muscular dystrophy. *Neurology* **60**, 993–997 (2003).
20. Senf, S. M., Dodd, S. L., McClung, J. M. & Judge, A. R. Hsp70 overexpression inhibits NF- κ B and Foxo3a transcriptional activities and prevents skeletal muscle atrophy. *FASEB J.* **22**, 3836–3845 (2008).
21. Park, H. S., Lee, J. S., Huh, S. H., Seo, J. S. & Choi, E. J. Hsp72 functions as a natural inhibitory protein of c-Jun N-terminal kinase. *EMBO J.* **20**, 446–456 (2001).
22. Chung, J. *et al.* HSP72 protects against obesity-induced insulin resistance. *Proc. Natl Acad. Sci. USA* **105**, 1739–1744 (2008).
23. Tupling, A. R. *et al.* HSP70 binds to the fast-twitch skeletal muscle sarco(endo)plasmic reticulum Ca^{2+} -ATPase (SERCA1a) and prevents thermal inactivation. *J. Biol. Chem.* **279**, 52382–52389 (2004).
24. Bornman, L., Polla, B. S., Lotz, B. P. & Gericke, G. S. Expression of heat-shock/stress proteins in Duchenne muscular dystrophy. *Muscle Nerve* **18**, 23–31 (1995).
25. Lou, J. S., Weiss, M. D. & Carter, G. T. Assessment and management of fatigue in neuromuscular disease. *Am. J. Hosp. Palliat. Care* **27**, 145–157 (2010).
26. Finsterer, J. Cardiopulmonary support in Duchenne muscular dystrophy. *Lung* **184**, 205–215 (2006).
27. Stedman, H. H. *et al.* The *mdx* mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy. *Nature* **352**, 536–539 (1991).
28. Briguet, A., Courdier-Fruh, I., Foster, M., Meier, T. & Magyar, J. P. Histological parameters for the quantitative assessment of muscular dystrophy in the *mdx*-mouse. *Neuromuscul. Disord.* **14**, 675–682 (2004).
29. Literati-Nagy, B. *et al.* Improvement of insulin sensitivity by a novel drug, BGP-15, in insulin-resistant patients: a proof of concept randomized double-blind clinical trial. *Horm. Metab. Res.* **41**, 374–380 (2009).
30. Morimoto, R. I. Cells in stress: transcriptional activation of heat shock genes. *Science* **259**, 1409–1410 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Koopman, J. G. Ryall and G. I. Lancaster for comments; J. Trieu, B. G. Gleeson, T. Naim and A. Chee for technical support; and C. Angelini and the Neuromuscular bank of tissues and DNA samples – Telethon Network of Genetic Biobanks for the provision of human muscle specimens. We thank N-Gene R&D Inc. USA for providing the BGP-15 compound. This study was supported in part by research grants from the National Health and Medical Research Council (NHMRC; project numbers 1009114 to G.S.L. and 472650 and 1004441 to M.A.F.), Association Française contre les Myopathies (France, to G.S.L.) and the Muscular Dystrophy Association (USA, to G.S.L.). M.A.F. is a Senior Principal Research Fellow of the NHMRC. A.P.R. was supported by a NHMRC Biomedical career Development Award. S.L. was supported by a postdoctoral fellowship from the Swiss National Science Foundation. S.M.G. was supported by a National Heart Foundation Postgraduate Scholarship (Australia). D.C.H. was supported by a National Heart Foundation Post-Doctoral Fellowship.

Author Contributions S.M.G., J.D.S., C.v.d.P., M.A.F. and G.S.L. conceived and designed the experiments. S.M.G., T.A.S., C.v.d.P., D.C.H. and J.E.C. performed the experiments. M.A.F. and K.E.D. facilitated experiments through the provision of mice and experimental compounds. A.P.R. and S.L. performed experiments on muscle samples from patients with DMD and from controls. S.M.G., D.C.H., M.A.F., C.v.d.P. and G.S.L. analysed the data. S.M.G. and G.S.L. wrote the manuscript. All authors checked for scientific content and contributed to the final drafting of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.S.L. (gsl@unimelb.edu.au).

METHODS

Animals. Female *mdx* mice were crossed with male mice expressing a rat inducible Hsp72 transgene under the control of a β -actin promoter³¹. F₁ generation males were mated with female *mdx* mice to yield an equal proportion of *mdx*^{TG(+)} and *mdx*^{TG(-)} littermate controls. WT^{TG(-)} and WT^{TG(+)} mice were generated as described previously²². Genotyping was performed by PCR using primers described previously³². C57BL/10 and *mdx* mice used for heat therapy and BGP-15 treatment studies were obtained from the Animal Resources Centre (Canning Vale, Western Australia). Dystrophic *mdx* mice were treated from 4 weeks to 9 weeks of age (5 weeks) or from 4 weeks to 16 weeks of age (12 weeks) with BGP-15 (15 mg kg⁻¹ per day, oral gavage; N-Gene R&D Inc.). BGP-15 is a pharmacological inducer of Hsp72, which has been shown to be safe and well tolerated in FDA-approved Phase II clinical trials for diabetes and insulin resistance^{29,33}. For heat therapy, the mice were anaesthetized and the core body temperature was raised to about 40 °C with an infrared thermometer for 30 min every fourth day; body temperature was monitored from the external auditory meatus as described previously³⁴. The *dko* mice were bred in the Biological Research Facility at the University of Melbourne⁶. The *dko* mice are utrophin-null on an *mdx* background and show a severe and progressive muscular dystrophy that more closely mimics the phenotypic characteristics of DMD^{7,8}. The *dko* mice were treated with BGP-15 from 3–4 weeks to 10 weeks of age. For the survival study, *dko* mice were treated from 3–4 weeks until death or humane killing in accordance with the Animal Ethics Committee of The University of Melbourne (AEC). All experiments were approved by the AEC and conducted in accordance with the Australian code of practice for the care and use of animals for scientific purposes as stipulated by the National Health and Medical Research Council (Australia). Male mice were used for all experiments.

Muscle functional analyses. Mice were anaesthetized with sodium pentobarbitone (Nembutal; 60 mg kg⁻¹ intraperitoneal; Sigma-Aldrich, New South Wales, Australia) such that they were unresponsive to tactile stimuli. Contractile properties of diaphragm muscle strips were assessed *in vitro* as described in detail previously³⁷. In brief, strips of diaphragm muscle were bathed in oxygenated Krebs solution at 25 °C in a custom-made organ bath. Preparations were stimulated by supramaximal 0.2-ms square-wave pulses of 450 ms duration delivered by means of platinum plate electrodes flanking both sides of the muscle. Contractile properties of the TA muscle were assessed *in situ*³⁵. In brief, TA muscles maintained at 37 °C were stimulated by supramaximal (14 V) 0.2-ms square-wave pulses of 350 ms duration delivered by means of two wire electrodes resting on the peroneal nerve. All muscles were adjusted to optimum muscle length (L_0), determined from maximum isometric twitch force (P_0). Maximum isometric tetanic force (P_{0t}) was recorded from the plateau of the frequency–force relationship, and normalized to muscle cross-sectional area (specific force; sP_0), for comparisons between groups where appropriate^{36–38}. The susceptibility of TA and diaphragm muscles to contraction-induced injury was determined from the protocols that we have described in detail previously^{36–38}. Isolated muscles were maximally activated to produce isometric force and then stretched to perform an eccentric contraction (at a velocity of $2L_f s^{-1}$) at progressively increasing magnitudes of stretch. Maximum isometric force was determined before each eccentric contraction³⁷.

Morphological analysis. Muscles were trimmed of tendons and adhering non-muscle tissue, mounted in embedding medium, frozen in liquid-nitrogen-cooled isopentane, and stored at –80 °C. Transverse muscle sections (5 μ m) were cryosectioned from the mid-belly of each muscle. Sections were stained with haematoxylin/eosin to reveal general muscle architecture. Cross-sectional area (CSA) and minimal Feret's diameter was assessed in about 500 fibres from each section of diaphragm muscle, with the use of Carl Zeiss software (Axiovision 4.6.3). The minimal Feret's diameter is the minimum distance between opposing parallel tangents of a muscle fibre³⁹, and the variance coefficient is a highly sensitive parameter used for detecting differences between dystrophic and otherwise healthy muscles³⁹. Muscle collagen content was assessed from Van Gieson's stained cross-sections that were quantified with Axiovision 4.6.3 software.

Enriched SR isolation. Mixed hindlimb muscles (quadriceps, gastrocnemius, extensor digitorum longus, soleus and plantaris) and diaphragm muscles were diluted in ice-cold homogenization buffer (250 mM sucrose, 5 mM HEPES pH 7.0, 0.2% NaN₃). Protease inhibitor cocktail (P-8340; Sigma-Aldrich, Castle Hill, New South Wales, Australia) was added immediately before homogenization at 5 μ l per 100 mg muscle wet weight. The muscles were minced on ice with surgical scissors and homogenized with a Polytron homogenizer (PT2100; Kinematica, Inc. Dispersing and Mixing Technology, New York) at a power setting of 21 for three 20-s bursts separated by 45-s breaks on ice. To obtain a purified and enriched SR membrane fraction, the homogenates were centrifuged at 5,500g for 10 min at 4 °C to remove insoluble material. The supernatants were harvested and centrifuged at 12,500g for 18 min at 4 °C. The pellet was discarded and the supernatant was

centrifuged for a second time at 12,500g for 18 min at 4 °C. Supernatants were transferred to ultracentrifuge tubes, which were balanced and centrifuged at 50,000g (24,200 r.p.m. on a T-890 fixed-angle rotor; Thermo Scientific Sorvall WX100 ultracentrifuge) for 1 h at 4 °C. Supernatants were discarded and pellets were resuspended in 5 ml of homogenization buffer containing 600 mM KCl and incubated on ice for 30 min. Samples were centrifuged at 15,000g for 10 min at 4 °C to pellet and remove mitochondria. Supernatants were centrifuged again at 50,000g for 1 h at 4 °C. The final pellet (enriched SR membrane vesicles) was resuspended in homogenization buffer. Protein content was determined in triplicate⁴⁰.

SERCA activity assay. Ca²⁺-dependent SERCA activity was assessed in isolated enriched SR vesicles and whole-muscle homogenates on the basis of the methods of Leberer and colleagues, as described previously⁴¹. For whole-muscle homogenates, muscles were surgically excised from anaesthetized mice and stored at –80 °C. Muscle samples (about 20–50 mg) were diluted in about 200 μ l of ice-cold homogenization buffer. Protease inhibitor cocktail (P-8340) was added immediately before use at a concentration of 5 μ l per 100 mg of muscle tissue. Muscles were homogenized with a hand-held glass homogenizer, and then centrifuged for 10 min at 5,500g at 4 °C. Supernatant protein concentration was determined in triplicate⁴⁰. Protein concentration was adjusted to 10 mg ml⁻¹, when possible, with homogenization buffer. SERCA activity in whole-muscle homogenates was determined in reaction buffer (200 mM KCl, 20 mM HEPES, pH 7.0, 15 mM MgCl₂, 10 mM NaN₃, 10 mM phosphoenolpyruvate, 5 mM ATP, 1 mM EGTA). SERCA activity in enriched SR vesicles was determined in reaction buffer (100 mM KCl, 20 mM HEPES, 10 mM MgCl₂, 10 mM NaN₃, 10 mM phosphoenolpyruvate, 5 mM ATP, and 1 mM EGTA). The pH of both reaction buffers was adjusted to 7.0 at 37 °C. Immediately before starting the reaction, 18 U ml⁻¹ PK, 18 U ml⁻¹ lactate dehydrogenase (LDH), 5 μ l NADH (100 mM), 1 μ M calcein A-23187 (Sigma-Aldrich) and about 10 μ l of whole-muscle homogenate or about 3 μ l of enriched SR vesicles were added to 1 ml of reaction buffer in a plastic cuvette. Cuvettes were loaded into a spectrophotometer and A_{340} was measured at 37 °C (Multiscan Spectrum; Thermo Electron, Waltham, Massachusetts). Maximal (V_{max}) and Ca²⁺-dependent SERCA activities were determined by progressively adding Ca²⁺ until a plateau or maximal activity was reached. The specific SERCA inhibitor 2',5'-di(tert-butyl)-1,4-benzohydroquinone (TBQ) was added to a final concentration of 40 μ M to determine basal activity. SERCA enzyme kinetic parameters, determined from a regression analysis, were the Ca²⁺ concentration required to elicit 50% maximal SERCA activity ($[Ca^{2+}]_{50}$) and the Hill coefficient, which is a measure of the cooperativity of Ca²⁺ binding of the SERCA enzyme. SERCA activity was graphed against added Ca²⁺ concentration. Non-regression analysis was performed using the following variable-slope sigmoidal dose–response relationship, using Graphpad Prism v.3.02 (GraphPad Software Inc., San Diego):

$$Y = Y_{bottom} + Y_{top} - \frac{Y_{bottom}}{1 + 10^{\log[Ca^{2+}]_{50} - X} n_H}$$

Peroxynterite-mediated SERCA inhibition assay. SERCA activity in enriched SR vesicles after incubation in various concentrations of peroxynterite ([ONOO⁻]) was measured. [ONOO⁻] was measured spectrophotometrically ($\epsilon = 1,670 \text{ M cm}^{-1}$ at 302 nm, about 68 mM in stock). Enriched SR vesicle samples were added to reaction buffer (as described previously) in the presence of ONOO⁻ (0, 10, 50, 100, 250, 500 or 1,000 μ M) and incubated for 5 min at 37 °C in a plastic cuvette. ONOO⁻ and byproducts were quenched by the addition of 1 mM dithiothreitol (to prevent the inactivation of PK and LDH subsequently added); 18 U ml⁻¹ PK, 18 U ml⁻¹ LDH and 5 μ l of NADH (100 mM) were then added. Cuvettes were loaded into a spectrophotometer, and A_{340} was measured at 37 °C. Maximal SERCA activity was determined by adding Ca²⁺ until a plateau or maximal activity was reached. Once maximal activity had been determined, the specific SERCA inhibitor TBQ was added to a final concentration of 40 μ M to determine basal activity. Maximal SERCA activity in each sample was assessed independently at various ONOO⁻ concentrations (10, 50, 100, 250, 500 or 1,000 μ M) and expressed as a percentage of the maximal SERCA activity in the absence of ONOO⁻.

Human DMD samples. Human muscle specimens were sourced from the Telethon Network of Genetic Biobanks, Italy. Biopsies were taken from the vastus lateralis muscle of patients with DMD (aged 1–9 years) or healthy controls (aged 18–25 years) using either a Bergström or open biopsy technique. All biopsies were frozen immediately in liquid nitrogen and stored at –80 °C until analysis. Muscle samples were homogenized in RIPA buffer (Millipore, Billerica, Massachusetts) and rotated for 1 h at 4 °C, followed by centrifugation for 15 min at 4 °C. The supernatant was collected and protein concentration was determined by means of the bicinchoninic acid protein assay (Pierce Biotechnology). Electrophoresis was performed by 10% SDS–PAGE in a buffer containing 12 mM Tris–HCl pH 8.8,

200 mM glycine and 0.1% SDS. Protein transfer was performed in a buffer containing 12 mM Tris-HCl pH 8.3, 200 mM glycine and 10% methanol with the use of poly(vinylidene difluoride) membranes. The membranes were blocked with 5% BSA in PBS, after which they were incubated overnight at 4 °C with the primary antibody against Hsp70 (SPA-812; Stressgen) diluted 1:1000 in PBS. After being washed, the membranes were incubated for 1 h with a goat anti-rabbit IgG antibody labelled with an infrared-fluorescent 800-nm dye (Alexa Fluor 800; Invitrogen) diluted 1:5,000 in PBS containing 50% Odyssey blocking buffer (LI-COR Biosciences) and 0.01% SDS. After being washed, the proteins were exposed on an Odyssey Infrared Imaging System (LI-COR Biosciences) and individual protein band optical densities were determined with ImageJ Software (National Institutes of Health). The blots were normalized against glyceraldehyde-3-phosphate dehydrogenase (GAPDH) protein (G8795; Sigma-Aldrich, Sydney, Australia).

Evans blue dye uptake. To quantify muscle damage and areas of focal necrosis, EBD (1% w/v) was injected intraperitoneally (10 µl per gram body mass). Muscles were frozen 20 h after the EBD injection. Sections 10 µm thick were cut on a cryostat, and EBD was detected as red autofluorescence with the use of a fluorescence microscope.

Wire hang test. A wire hang test was employed to assess whole-body muscle strength and endurance. Mice were placed on a wire mesh grid, which they gripped; the grid was then inverted. Latency-to-fall onto a padded surface was recorded in three successive trials, with the average of the best two trials used for analyses.

Single muscle fibre analysis. Mice were anaesthetized with sodium pentobarbitone (Nembutal) such that they were unresponsive to tactile stimuli, and the extensor digitorum longus and/or soleus muscle was surgically excised for single fibre analysis. The muscle was blotted on filter paper and placed in a Petri dish containing paraffin oil at room temperature (23 ± 2 °C). Muscles were pinned at resting length to the base of a dish. Single muscle fibres were isolated from as close to the surface of the muscle as possible, and fine forceps were used to peel the sarcolemma away from the contractile apparatus under a dissecting microscope. The mechanically skinned fibre was then attached to one end of a piezoresistive force transducer (AE801 SensoNor; Horten) using braided silk (size 10, 0.2 mm; Deknatel), and the other end of the fibre was clamped between a pair of forceps fixed to a micromanipulator. All experiments were conducted at room temperature. All solutions had pH of 7.10 ± 0.01 , and the free Mg^{2+} concentration ($[Mg^{2+}]$) was 1 mM, unless specified otherwise. Free $[Ca^{2+}]$ at 0.1 µM or more was verified with a Ca^{2+} -sensitive electrode (Orion Research).

Caffeine-induced force responses and SR Ca^{2+} accumulation. Initially, mechanically skinned muscle fibres were equilibrated for 30 s in a wash solution followed by thorough depletion of SR Ca^{2+} stores, achieved by transferring the fibre preparation into a release solution containing 30 mM caffeine and 0.02 mM free Mg^{2+} . The presence of 0.5 mM EGTA in the release solution ensured that the level of Ca^{2+} during caffeine-induced release did not maximally activate the contractile apparatus, which is necessary to allow quantitative evaluation of the amount of Ca^{2+} released. Ca^{2+} release from the SR was estimated from the relative areas under the caffeine-induced force response. The fibre was left in the release solution for 2 min to ensure complete SR Ca^{2+} depletion, before being washed for 30 s. Thereafter, the SR was reloaded with Ca^{2+} in load solution (0.2 µM Ca^{2+} (pCa 6.7), where $pCa = -\log_{10}[Ca^{2+}]$) for increasing durations (10, 20, 30 and 60 s), before being equilibrated for 30 s in wash solution; subsequently, SR Ca^{2+} was released in release solution. Data were fitted with a standard exponential association equation to give the rate at which the SR accumulated Ca^{2+} (in s^{-1}) but not the amount of SR Ca^{2+} accumulated. SR accumulation is an indirect measure of SERCA activity⁴².

SR Ca^{2+} leak. The percentage of Ca^{2+} lost from the SR as a result of the passive leak was also assessed. The fibre was loaded for 20 s in loading solution. The fibre preparation was then placed in wash solution for 30 s followed by SR Ca^{2+} content released in release solution (Ca^{2+} leak in 30 s). The fibre preparation was placed in wash solution before reloading for 20 s in load solution and transferred to wash solution for 90 s; the remaining SR Ca^{2+} was released in release solution (Ca^{2+} leak in 90 s). The 30-s Ca^{2+} leak was then repeated, and the area (corrected for proportionality between area and SR Ca^{2+} content) under the test run was divided by the average of the areas under the caffeine-induced force responses in the controls before and after the test run. This gave an estimate of the fraction of SR Ca^{2+} leaked over a 60-s leak period.

Relative SR Ca^{2+} sensitivity. To determine the effect of Hsp72 overexpression on the ryanodine receptor (RyR), a caffeine dose-response curve was determined from the forces produced by the contractile apparatus after SR Ca^{2+} release induced by low caffeine concentrations. Each fibre was prepared by completely depleting the SR of Ca^{2+} with 30 mM caffeine followed by a 30-s SR Ca^{2+} reloading duration. The peak force of caffeine-induced contraction was determined in a

series of potassium hexamethylene-diamine tetraacetic acid solutions containing 2, 3, 5 and 7 mM caffeine and 50 µM EGTA, with complete SR Ca^{2+} depletion and 30-s reloading duration of the SR with Ca^{2+} between each caffeine contraction. The peaks of the caffeine-induced contractions were then normalized as a percentage of maximum Ca^{2+} -activated force to estimate the sensitivity of the RyR to caffeine. This determined whether overexpression of Hsp72 directly affected the function of the RyR.

Properties of the contractile apparatus. After SR properties were investigated, the single muscle fibres were equilibrated in a relaxing solution (pCa > 9) for 2 min. Fibres were placed in a maximum Ca^{2+} -activating solution (pCa ≈ 4.5) until force reached the maximal value (maximum Ca^{2+} -activated force) and then placed back in the relaxing solution for a further 2 min. Force responses were generated by exposing the fibre to activating solutions of progressively lower pCa (higher $[Ca^{2+}]$) in a stepwise fashion. The force response generated at each pCa was expressed as a percentage of the interpolated values for maximum Ca^{2+} -activated force. Data points were fitted with a Hill equation producing two parameters: the pCa_{50} (the pCa producing half-maximum force) and n_H (the Hill coefficient, indicative of the steepness of the force-pCa relationship).

Western blotting. After the determination of SERCA activity in whole-muscle homogenates and enriched SR vesicles, electrophoresis was performed on the same samples for the quantification of protein expression. Equal amounts of protein were resolved in SDS buffer, heated for 5 min at 95 °C and separated on SDS-polyacrylamide gels. Separated proteins were transferred to poly(vinylidene difluoride) membranes (0.45-µm Immobilon-P; Millipore). Membranes were blocked with 5% non-fat milk powder (or BSA) in Tris-buffered saline containing Tween 20 for 1 h and incubated overnight at 4 °C with appropriate antibody dilutions. Antibodies against SERCA 1 (Affinity Bioreagents), SERCA 2a (Affinity Bioreagents), Hsp72 (Stressgen), α -tubulin (ECM-Biosciences) and GAPDH (Sigma) were used. Membranes were digitized with a chemiluminescent detection and imaging system (ChemiDoc XRS; Bio-Rad Laboratories, Hercules, California, USA) and band densities were quantified with Quantity One analysis software (version 4.6.1; Bio-Rad Laboratories).

Real-time PCR. Real-time PCR was performed as described previously⁴³. Each sample was run in triplicate. The mean reading of each triplicate was converted to an absolute content by using a DNA standard curve, based on a serial dilution of DNA standard (100–10,000 ng ml⁻¹; Molecular Probes) run together with the samples on each plate. Gene expression was quantified by normalizing the logarithmic cycle threshold (C_T) value ($2C_T$) to the cDNA content of each sample to obtain the expression $2C_T/cDNA$ content (ng ml⁻¹). In Supplementary Fig. 5b, mRNA for SERCA 1 and SERCA 2a was normalized to the eukaryotic 18S house-keeping gene.

Creatine kinase analysis. Serum CK activity was analysed as an overall measure of whole-body muscle breakdown. Blood was collected from the tail vein and centrifuged for 10 min at 10,000g and 4 °C to isolate the serum fraction. Serum CK activity was then determined with a commercially available creatine kinase assay kit (ECPK-100) in accordance with the manufacturer's instructions (BioAssay Systems).

Skeletal preparation. Skeletal architecture was revealed as described previously⁴⁴. After death, the mice were skinned, eviscerated and placed in a KOH solution (1.5% w/v), for 5 days. KOH solution was replaced and a small amount of alizarin red was added to stain calcium deposits; the preparation was left for a further 5 days to produce the specimens shown in Fig. 4a.

Statistical analysis. All values are presented as means \pm s.e.m. Unpaired Student's *t*-tests were used to compare between two groups. When comparisons were being made between more than two groups, a one-way analysis of variance was used with Newman-Keuls post-hoc multiple comparison test.

- Marber, M. S. *et al.* Overexpression of the rat inducible 70-kD heat stress protein in a transgenic mouse increases the resistance of the heart to ischemic injury. *J. Clin. Invest.* **95**, 1446–1456 (1995).
- Taleb, M. *et al.* Hsp70 inhibits aminoglycoside-induced hair cell death and is necessary for the protective effect of heat shock. *J. Assoc. Res. Otolaryngol.* **9**, 277–289 (2008).
- Literati-Nagy, B. *et al.* Beneficial effect of the insulin sensitizer (HSP inducer) BGP-15 on olanzapine-induced metabolic disorders. *Brain Res. Bull.* **20**, 340–344 (2010).
- Saegusa, Y. & Tabata, H. Usefulness of infrared thermometry in determining body temperature in mice. *J. Vet. Med. Sci.* **65**, 1365–1367 (2003).
- Gregorevic, P., Plant, D. R., Leeding, K. S., Bach, L. A. & Lynch, G. S. Improved contractile function of the *mdx* dystrophic mouse diaphragm muscle after insulin-like growth factor-I administration. *Am. J. Pathol.* **161**, 2263–2272 (2002).
- Schertzer, J. D., Ryall, J. G. & Lynch, G. S. Systemic administration of IGF-I enhances oxidative status and reduces contraction-induced injury in skeletal muscles of *mdx* dystrophic mice. *Am. J. Physiol.* **291**, 499–505 (2006).
- Schertzer, J. D., Gehrig, S. M., Ryall, J. G. & Lynch, G. S. Modulation of insulin-like growth factor (IGF)-I and IGF-binding protein interactions enhances skeletal

- muscle regeneration and ameliorates the dystrophic pathology in *mdx* mice. *Am. J. Pathol.* **171**, 1180–1188 (2007).
38. Gehrig, S. M., Koopman, R., Naim, T., Tjoakarfa, C. & Lynch, G. S. Making fast-twitch dystrophic muscles bigger protects them from contraction injury and attenuates the dystrophic pathology. *Am. J. Pathol.* **176**, 29–33 (2010).
 39. Brigue, A., Courdier-Fruh, I., Foster, M., Meier, T. & Magyar, J. P. Histological parameters for the quantitative assessment of muscular dystrophy in the *mdx*-mouse. *Neuromuscul. Disord.* **14**, 675–682 (2004).
 40. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
 41. Schertzer, J. D. *et al.* β_2 -Agonist administration increases sarcoplasmic reticulum Ca^{2+} -ATPase activity in aged rat skeletal muscle. *Am. J. Physiol.* **288**, 526–533 (2005).
 42. Plant, D. R. & Lynch, G. S. Depolarization-induced contraction and SR function in mechanically skinned muscle fibres from dystrophic *mdx* mice. *Am. J. Physiol.* **285**, C522–C528 (2003).
 43. Murphy, K. T. *et al.* Antibody-directed myostatin inhibition in 21-mo-old mice reveals novel roles for myostatin signaling in skeletal muscle structure and function. *FASEB J.* **24**, 4433–4442 (2010).
 44. Green, M. C. A rapid method for clearing and staining specimens for the demonstration of bone. *Ohio J. Sci.* **52**, 31–33 (1952).

Crystal structure of a membrane-embedded H^+ -translocating pyrophosphatase

Shih-Ming Lin¹, Jia-Yin Tsai², Chwan-Deng Hsiao², Yun-Tzu Huang¹, Chen-Liang Chiu¹, Mu-Hsuan Liu¹, Jung-Yu Tung¹, Tseng-Huang Liu¹, Rong-Long Pan¹ & Yuh-Ju Sun¹

H^+ -translocating pyrophosphatases (H^+ -PPases) are active proton transporters that establish a proton gradient across the endomembrane by means of pyrophosphate (PP_i) hydrolysis^{1,2}. H^+ -PPases are found primarily as homodimers in the vacuolar membrane of plants and the plasma membrane of several protozoa and prokaryotes^{2,3}. The three-dimensional structure and detailed mechanisms underlying the enzymatic and proton translocation reactions of H^+ -PPases are unclear. Here we report the crystal structure of a *Vigna radiata* H^+ -PPase (VrH^+ -PPase) in complex with a non-hydrolysable substrate analogue, imidodiphosphate (IDP), at 2.35 Å resolution. Each VrH^+ -PPase subunit consists of an integral membrane domain formed by 16 transmembrane helices. IDP is bound in the cytosolic region of each subunit and trapped by numerous charged residues and five Mg^{2+} ions. A previously undescribed proton translocation pathway is formed by six core transmembrane helices. Proton pumping can be initialized by PP_i hydrolysis, and H^+ is then transported into the vacuolar lumen through a pathway consisting of Arg 242, Asp 294, Lys 742 and Glu 301. We propose a working model of the mechanism for the coupling between proton pumping and PP_i hydrolysis by H^+ -PPases.

Two proton-pumping proteins, vacuolar H^+ -ATPases (V-ATPases) and H^+ -PPases, coexist on plant vacuolar membranes and use ATP and PP_i , respectively, as energy sources for H^+ translocation⁴. Both of these enzymes acidify the vacuolar lumen and establish an electrochemical proton gradient across the vacuolar membrane. Whereas V-ATPases have been widely studied, the structure and function of H^+ -PPases are as yet unknown. The H^+ -PPase was first identified from photosynthetic bacterium *Rhodospirillum rubrum*⁵, after which the functional and biochemical investigations of vacuolar H^+ -PPases from plants were reported^{6,7}. Subsequently, the protein chemical identification of H^+ -PPases from red beet⁸ and mung bean⁹ and the molecular cloning of the enzyme from *Arabidopsis*¹⁰ were gradually completed. Afterwards, the heterologous expression of H^+ -PPase in yeast was established and the sufficiency of PP_i -energized H^+ translocation was demonstrated¹¹. On the basis of these arduous results, a fourth category of primary H^+ pump, H^+ -PPases, was thus determined¹².

H^+ -PPases have a high degree of amino acid sequence homology (86–91% identity in land plants; Supplementary Fig. 1)², and they can be divided into two subfamilies: type I (K^+ -dependent) and type II (K^+ -independent)¹³. In *Arabidopsis*, the H^+ -PPase regulates organ development by modulating the apoplastic pH and eliminating the metabolic by-product, cytosolic PP_i (refs 14, 15). The overexpression of H^+ -PPases improves tolerance to salinity and drought in many higher plants by increasing ion retention^{16,17}. A PP_i -binding motif, (E/D)(X)₇KXE, and two acidic motifs, DX₃DX₃D, have been suggested as being essential for the enzymatic function of H^+ -PPases^{2,12,18}. Nevertheless, the three-dimensional structure of H^+ -PPase is unknown. We therefore sought to solve its atomic structure to gain a better understanding of the machinery of this unique proton pump.

We determined the crystal structure of VrH^+ -PPase bound to IDP by the multiple-wavelength anomalous dispersion (MAD) method and multiple isomorphous replacement with anomalous scattering (MIRAS) (Supplementary Table 1). The protein is compactly folded in a rosette manner with 16 transmembrane helices (TMs M1–M16) within two concentric walls (Fig. 1a). Six TMs (M5, M6, M11, M12, M15 and M16) bundle together to form an inner wall, which is in turn surrounded by ten additional TMs (M1–M4, M7–M10 and M13–M14) that constitute the outer wall (Fig. 1b). The TMs fold sequentially counterclockwise in the inner and outer walls with the exception of M1, which tilts slightly outward. There are two short helices ($\alpha 2$ and $\alpha 3$) on the cytosolic side, and two additional helices ($\alpha 1$ and $\alpha 4$) together with two antiparallel β strands ($\beta 1$ and $\beta 2$) on the luminal side (Supplementary Fig. 2). A disulphide bond (C124–C132) was detected in loop 2 (Supplementary Fig. 2). Both the amino and carboxy termini face the vacuolar lumen (Fig. 1c). Each subunit contains a single IDP molecule embedded in the core of the inner wall near the cytosolic region. It is conceivable that the inner wall is responsible for substrate binding, and the outer wall probably maintains the structural integrity of the protein.

VrH^+ -PPase is a homodimer with an extensively buried 3,241 Å² surface between the subunits (Fig. 1c). The electrostatic surface potential of the VrH^+ -PPase dimer is shown in Fig. 1d. The cytosolic region of VrH^+ -PPase has many hydrophilic residues, whereas the vacuolar region of the protein protruding out of the membrane is smaller. The homodimer has two-fold symmetry and a root mean squared deviation of 0.32 Å among the C α atoms between the two subunits (Supplementary Fig. 3a). Four TMs (M5, M12, M15 and M16) from the inner wall and two TMs (M10 and M13) from the outer wall participate in subunit interactions that primarily include hydrophobic interactions, six hydrogen bonds and two salt bridges (Supplementary Fig. 3b–d and Supplementary Table 2).

The substrate/IDP-binding site is a funnel-shaped pocket formed by six core TMs with a solvent-accessible volume of 1,521 Å³ (Fig. 2a). The electrostatic surface potential of the IDP-binding pocket is shown in Fig. 2b. The pocket has an unusually acidic environment that contains 12 acidic residues (Asp 253, Asp 257, Glu 268, Asp 269, Asp 279, Asp 283, Asp 287, Asp 507, Asp 691, Asp 723, Asp 727 and Asp 731). Three lysine residues (Lys 250, Lys 730 and Lys 694) and one asparagine residue (Asn 534) are also found in this pocket (Fig. 2c). All these residues are highly conserved among H^+ -PPases (Supplementary Fig. 1). Site-directed mutagenesis studies indicate that most of these conserved residues are essential for PP_i hydrolysis^{18–20}. In the current structure, the IDP molecule was bound directly by three lysine residues through hydrogen bonds. In addition, one K^+ ion and five Mg^{2+} ions (Mg1–Mg5) were identified around the IDP molecule at the pocket (Fig. 2c). K^+ is the stimulator of the type I H^+ -PPases¹, and Mg^{2+} is essential for the activity of H^+ -PPases²¹. All these Mg^{2+} ions mediate the interactions between IDP and Asp/Asn residues. The binding interactions (Supplementary Table 3) precisely confine the

¹Department of Life Science and Institute of Bioinformatics and Structural Biology, College of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan. ²Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan.

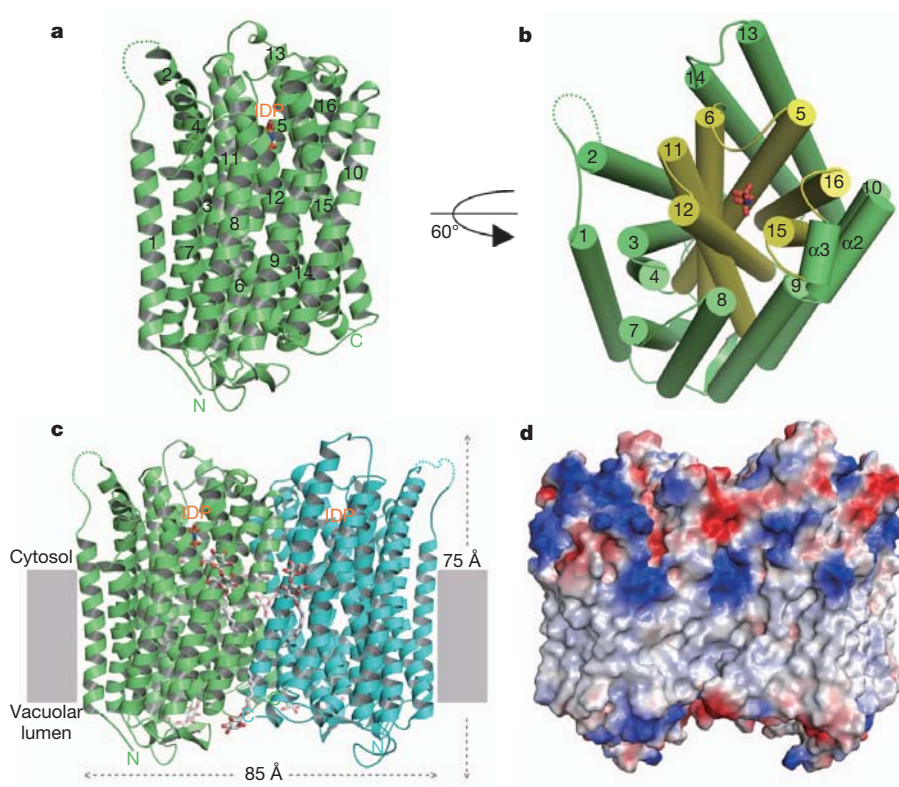


Figure 1 | The crystal structure of the V_{rh}^{+} -PPase-IDP complex. **a**, Ribbon diagram of the overall structure of V_{rh}^{+} -PPase, containing 16 TMs (labelled 1–16). A missing region (residues 42–66) is shown with dotted lines. **b**, The six inner and ten outer TMs drawn as cylinders and coloured in yellow and green, respectively. This orientation is rotated by 60° from that in **a**. **c**, V_{rh}^{+} -PPase

dimer shown as a ribbon diagram with height and width dimensions of 75 Å and 85 Å, respectively. The detergent molecules of n-decyl β -D-maltoside are shown as sticks. **d**, Electrostatic surface potential of the V_{rh}^{+} -PPase dimer (red, blue and white indicate negative, positive and neutral potentials, respectively). In **a**–**c**, IDP is shown as sticks and coloured in CPK.

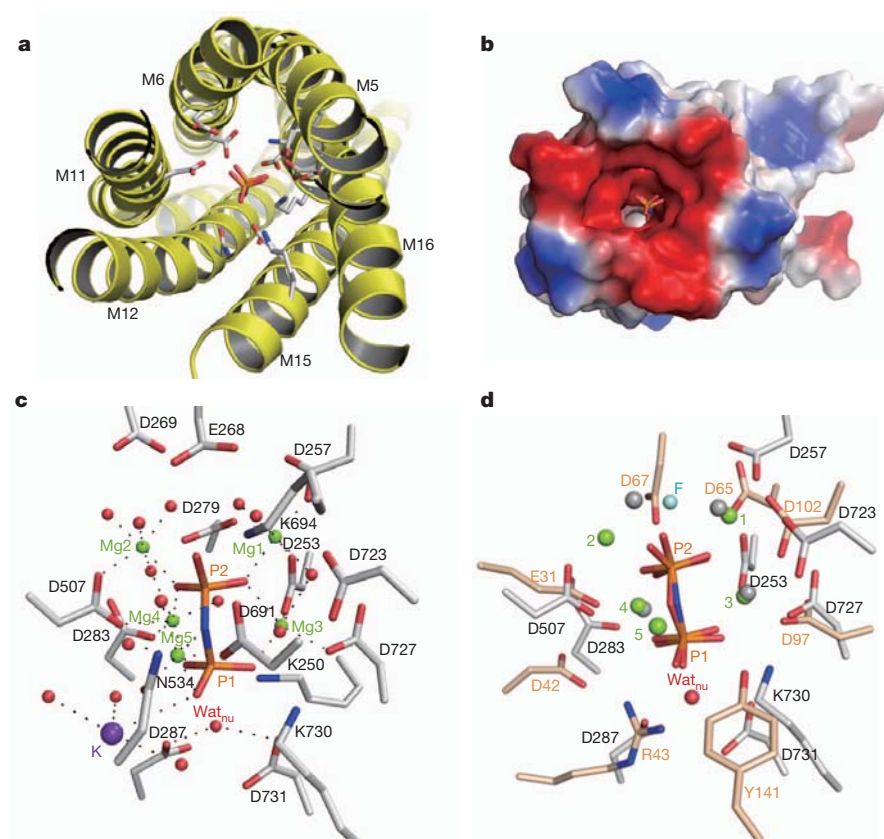


Figure 2 | The substrate-binding site. **a**, Six core TMs (yellow ribbon) with IDP-binding residues (sticks). **b**, The electrostatic surface potential of the IDP binding pocket (red, blue and white indicate negative, positive and neutral potentials, respectively). **c**, The IDP-binding residues. Mg^{2+} ions, K^{+} ions and water molecules are shown as green, purple and red spheres, respectively. Interactions are presented as dashed lines. **d**, The binding site of V_{rh}^{+} -PPase (in white) superimposed on *Ec*PPase (PDB 2AUU; in pink). The Mg^{2+} ions of the V_{rh}^{+} -PPase (in green, with numbers) and *Ec*PPase (in grey) are shown as spheres. The F^{-} in *Ec*PPase is shown as a blue sphere, and the Wat_{nu} in V_{rh}^{+} -PPase is shown as a labelled red sphere. IDP in V_{rh}^{+} -PPase and PP_i in *Ec*PPase are coloured in CPK.

substrate at the active site in a proper orientation for enzymatic hydrolysis.

The current structure indicates that the highly conserved fragment D253-X₃-D257-X₃-K261-X-E263 from M5, corresponding to the DX₇KXE motif, is involved in substrate binding^{12,18}. Both Asp 253 and Asp 257 interact with IDP through Mg1 and Mg3 (Fig. 2c). However, Lys 261 and Glu 263 do not participate directly in IDP binding. Lys 261 contributes to a salt-bridge network (Supplementary Fig. 4 and Supplementary Table 4) that connects the core TMs to stabilize the active site. Glu 263 forms a salt bridge with Arg 609 and probably contributes to the structural stability of V_rH⁺-PPase. Furthermore, the highly conserved acidic motifs D279-X₃-D283-X₃-D287 and D723-X₃-D727-X₃-D731 as well as all aspartic acids, except Asp 287 and Asp 731, interact with IDP through either Mg²⁺ ions or water.

In Fig. 2d the substrate and the surrounding residues in V_rH⁺-PPase are superimposed on those of type I soluble *Escherichia coli* pyrophosphatase (EcPPase)²². Both enzymes contain several aspartic residues for coordinating Mg²⁺ ions to bind the substrate at the active site. There are five and four Mg²⁺ ions in the binding pockets of V_rH⁺-PPase and EcPPase, respectively. The binding of Mg5 to IDP is detected only in V_rH⁺-PPase. The locations of Mg3 and Mg4 in V_rH⁺-PPase are similar to those of the corresponding Mg²⁺ ions in EcPPase. In contrast, Mg1 and Mg2 are located at distinct positions in both enzymes. The distances between Mg1 and Mg2 in V_rH⁺-PPase and EcPPase are 5.3 and 3.5 Å, respectively. In EcPPase, Mg1 and Mg2 bind with a specific inhibitor, F⁻, which occupies the position of the nucleophile close to the P2 phosphate of PP_i, to prevent hydrolysis of the substrate²². However, a similar binding phenomenon is not observed in V_rH⁺-PPase, because the distance between Mg1 and Mg2 is too large to bind a nucleophile for attacking PP_i (Fig. 2d). Instead, in V_rH⁺-PPase, a water molecule (Wat_{nu}) is found near the P1 phosphate of IDP (2.6 Å) and forms hydrogen bonds with Asp 287 and Asp 731, respectively. It is conceivable that Wat_{nu} might act as a nucleophile for PP_i hydrolysis in V_rH⁺-PPase. Asp 287 and Asp 731 are highly conserved among H⁺-PPases but are replaced by Arg 43 and Tyr 141 in soluble EcPPase. F⁻ cannot compete with Wat_{nu} for the same binding site, presumably as a result of the repulsion between negatively charged F⁻ and aspartic acid. H⁺-PPase is therefore less sensitive to F⁻ inhibition than soluble PPase is². These differences suggest that membrane-bound H⁺-PPases and soluble PPases use different strategies to trap the nucleophile for PP_i hydrolysis.

Previous evidence suggests that the dimerization of H⁺-PPase is important for H⁺ translocation activity²³. However, the possible residues that can be protonated as part of proton conduction could not be found in the dimer interface in V_rH⁺-PPase. Instead, four lined-up charged residues, Arg 242, Asp 294, Lys 742 and Glu 301, were observed in a narrow, compact and water-inaccessible transmembrane region of the core TMs (Fig. 3a). These residues are highly conserved among H⁺-PPases except Glu 301, which is found only in higher plants (Supplementary Fig. 1). Glu 301 is located at a rather narrow point in the pathway and close to the vacuolar lumen. It might therefore act as a constricting neck (Fig. 3a). Mutations at Glu 301 abolish proton translocation activity (Supplementary Table 5)^{19,24,25}. In addition, Glu 301 contributes to the binding of the lipophilic carbodiimide *N,N'*-dicyclohexylcarbodiimide (DCCD), a potent blocker of proton conductance²⁴. These data indicate that Glu 301 probably serves as the proton donor/acceptor in the proton translocation mechanism. Furthermore, Asp 294 and Lys 742 form the only salt bridge among these acid–base pairs (Fig. 3a). Lys 742 might reciprocally modulate the protonation and deprotonation of Asp 294 and Glu 301, similar to that in the regulatory machinery of bacteriorhodopsin²⁶. Arg 242, which is located near the PP_i-binding site, may form a salt bridge with Asp 287 or Asp 731, resulting in the deprotonation of the latter residues. Site-directed mutagenesis studies revealed that Arg 242, Asp 294 and Lys 742 are crucial for PP_i hydrolysis as well as H⁺ translocation (Supplementary Table 5)^{19,27}. Together, these acid–base pairs,

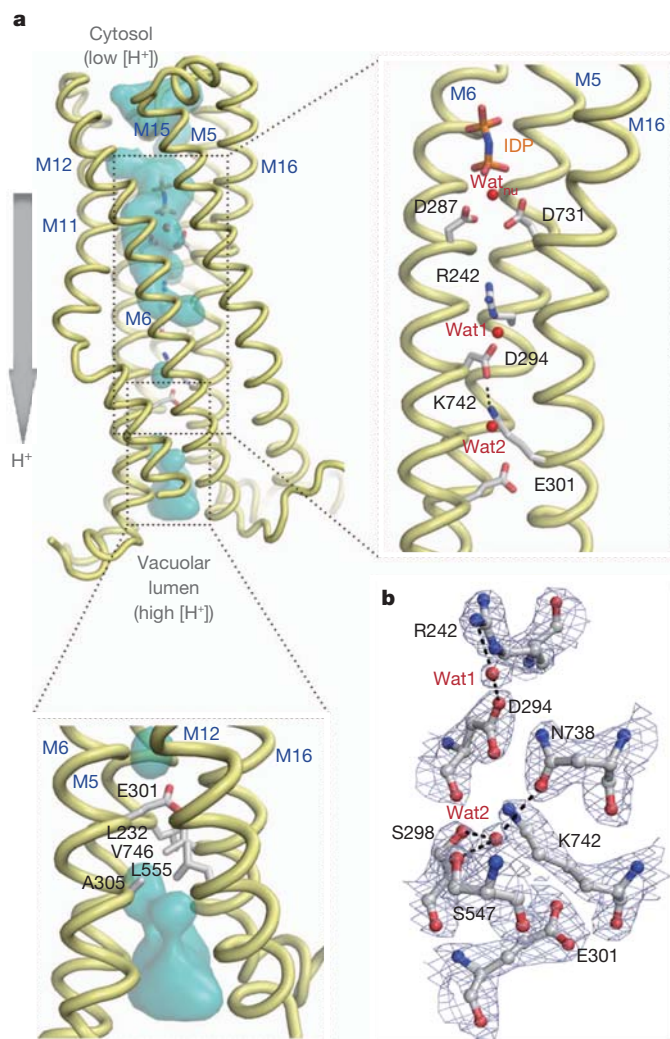


Figure 3 | The proton transport pathway of V_rH⁺-PPase. **a**, The proton transport pathway is formed by six core TMs (M5, M6, M11, M12, M15 and M16). The solvent-accessible surface area is coloured in cyan. The arrow indicates the direction of proton translocation. Right: zoomed-in view of the proton transport pathway. The residues involved in proton transport are shown and labelled. The solvent-accessible surface has been removed. Bottom: the hydrophobic gate around Glu 301. Residues forming a hydrophobic gate are displayed and labelled. **b**, The electron density map ($2F_{\text{obs}} - F_{\text{calc}}$) (in blue) around the proton transport pathway drawn at a contour level of 2σ . The IDP is shown as sticks and coloured in CPK. Water molecules Wat_{nu}, Wat1 and Wat2 are presented as labelled red spheres. Water-mediated hydrogen bonds are drawn as black dashed lines.

Arg 242, Asp 294, Lys 742 and Glu 301, form a potential proton transport pathway surrounded by six core TMs that convey protons from the active site to the vacuolar lumen.

The proton-pumping proteins usually use bound water molecules to assist proton transport²⁸; examples are bacteriorhodopsin²⁶ and P-type H⁺-ATPases²⁹. Nevertheless, two ordered water molecules (Wat1 and Wat2) along the proton transport pathway of V_rH⁺-PPase were identified from the electron density map (Fig. 3b). Wat1 forms hydrogen bonds with Arg 242 and Asp 294 and seems to be a continuation of the bulk solvent at the PP_i-binding site. Wat2 was found in the vicinity of Lys 742 (with a distance of 3.7 Å) and trapped by Asp 294, Ser 298, Ser 547, Asn 738 and Lys 742 with hydrogen bonding to Ser 298, Ser 547 and Asn 738. The bound waters, Wat1 and Wat2, presumably facilitate proton transport in the pathway of V_rH⁺-PPase.

In the V_rH⁺-PPase-IDP complex, the vacuolar portion of the proton transport pathway is relatively narrow and is occupied by hydrophobic residues, such as Leu 232 (M5), Ala 305 (M6), Leu 555 (M12) and

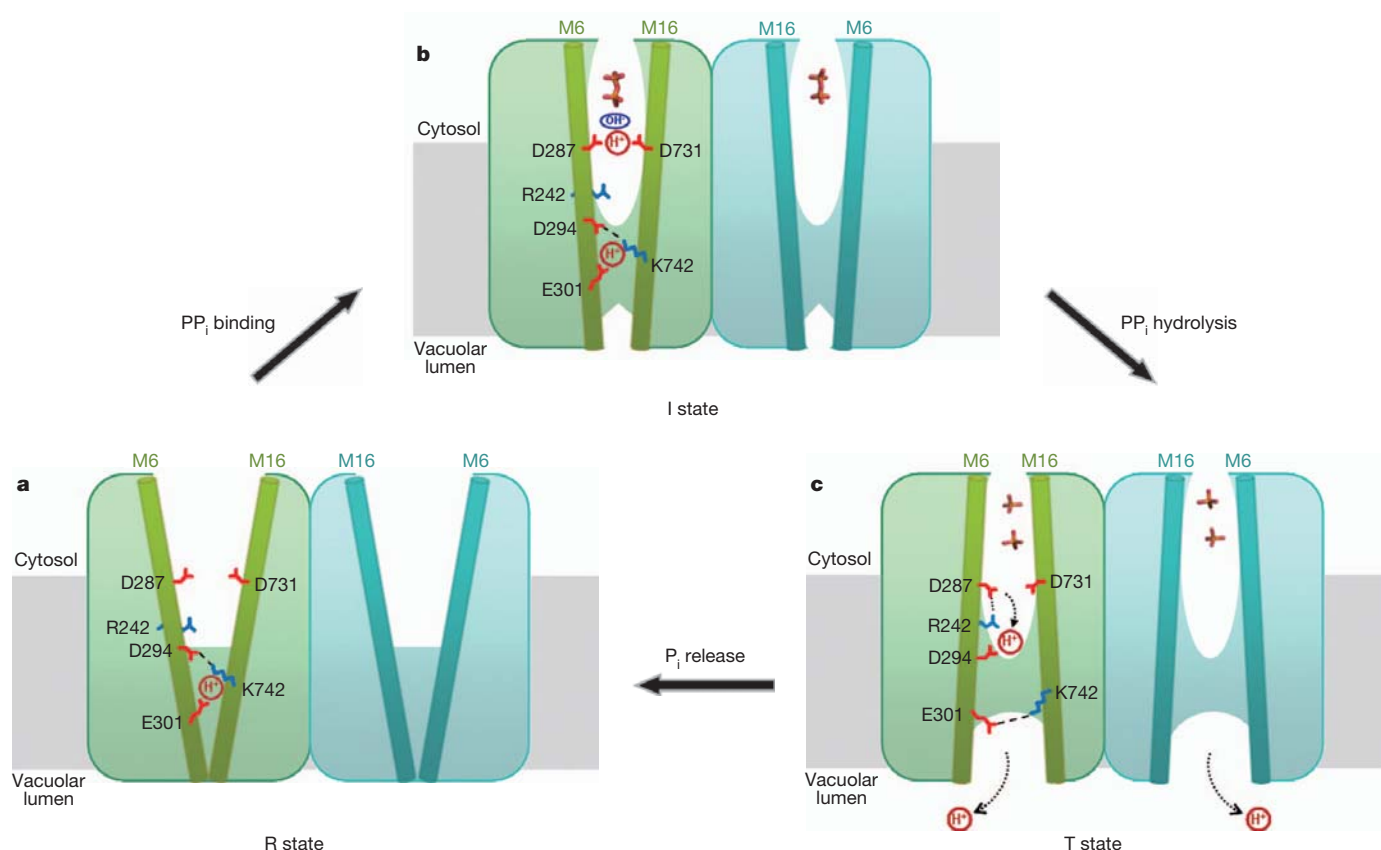


Figure 4 | A working model for proton pumping in VrH^+ -PPase. **a**, Resting state (R state). **b**, Initiated state (I state). **c**, Transient state (T state). The VrH^+ -PPase dimer is shown and coloured in green and blue. M6 and M16 are shown

Val 746 (M16) (Fig. 3a). Most of these residues are conserved among H^+ -PPases from various species (Supplementary Fig. 1). They possibly form a hydrophobic gate, keeping Glu 301 away from the hydrophilic environment of the vacuolar lumen, where the H^+ concentration is higher. Such a gate prevents the H^+ from refluxing and maintains unidirectional H^+ translocation from cytosol to lumen. It is conceivable that the constricted pathway and the acid–base pairs have crucial functions in directional proton pumping of the H^+ -PPase.

On the basis of previous mutational and biochemical studies (Supplementary Table 5) and the present structural findings, we propose a working model for the coupling of PP_i hydrolysis to proton pumping (Fig. 4). In this model, the process involves three sequential states: R, I and T. Similarly, the limited trypsinolysis analysis of H^+ -PPase^{20,21} (Supplementary Fig. 5) indicates that the binding pocket of H^+ -PPase is flexible and might be exposed to the solvent in the absence of the substrate, and the protein is in a resting state (R state; Fig. 4a). In addition, an open conformation at the active site in the absence of substrate (PP_i) was also supported by a study using single-molecule fluorescence resonance energy transfer (FRET)³⁰. In contrast, the luminal portion of the core TMs assumes a closed conformation that prevents H^+ from back-flushing into the cytosol. On binding of substrate (or IDP in our structure), H^+ -PPase enters an initiated state for PP_i hydrolysis (I state; Fig. 4b). The core TMs then change into a closed conformation on the cytosolic side, locking PP_i in the substrate-binding pocket. At this stage, the luminal portion of these TMs is changed into a semi-open conformation for sequential H^+ translocation. The structure of the VrH^+ -PPase-IDP complex that we have resolved would represent the protein in the I state. On hydrolysis of PP_i , free phosphates are generated with concomitant proton production (Fig. 4c). Then the proton relay through Arg 242–Asp 294–Lys 742–Glu 301 occurs as a chain reaction, and the core TMs adopt

as cylinders. The residues involved in proton transport are shown and labelled. PP_i and free P_i are shown as sticks and coloured in CPK. The proton and hydroxyl ions are labelled. The possible salt bridges are shown as dashed lines.

an open conformation on the vacuolar lumen side for proton release. Consequently, H^+ -PPase proceeds to the transient state (T state; Fig. 4c). After the release of P_i to the cytosol and translocation of a proton to the vacuolar lumen, H^+ -PPase returns to the R state. This proton-pumping mechanism is completed through a series of delicate conformational changes driven primarily by the energy derived from PP_i hydrolysis. Thus, our working model provides a scheme capable of accounting for how PP_i hydrolysis and the movement of protons from one side of the membrane to the other might be accomplished by the H^+ -PPase.

METHODS SUMMARY

Vigna radiata H^+ -translocating pyrophosphatase (VrH^+ -PPase) fused with a C-terminal His-tag was expressed and isolated from *Saccharomyces cerevisiae* with methods that have been described previously. In brief, VrH^+ -PPase was crystallized by using 23% (w/v) PEG2000 precipitant in 50 mM MES pH 6.5. The structural phase was determined from the tantalum derivative Ta_6Br_{12} by MAD and also from OsO_4 and Pt-organic derivatives by MIRAS (Supplementary Table 1). Model building was performed with COOT, and the final model was refined with REFMAC5 to an R -factor of 16.80% and an R_{free} of 20.31% at 2.35 Å. The refinement statistics are summarized in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 September 2011; accepted 17 February 2012.

Published online 28 March 2012.

1. Rea, P. A. & Poole, R. J. Vacuolar H^+ -translocating pyrophosphatase. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 157–180 (1993).
2. Maeshima, M. Vacuolar H^+ -pyrophosphatase. *Biochim. Biophys. Acta* **1465**, 37–51 (2000).
3. Serrano, A., Perez-Castineira, J. R., Baltscheffsky, H. & Baltscheffsky, M. Proton-pumping inorganic pyrophosphatases in some archaea and other extremophilic prokaryotes. *J. Bioenerg. Biomembr.* **36**, 127–133 (2004).

4. Maeshima, M. Tonoplast transporters: organization and function. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 469–497 (2001).
5. Baltscheffsky, H., Von Stedingk, L. V., Heldt, H. W. & Klingenberg, M. Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science* **153**, 1120–1122 (1966).
6. Chanson, A., Fichmann, J., Spear, D. & Taiz, L. Pyrophosphate-driven proton transport by microsomal membranes of corn coleoptiles. *Plant Physiol.* **79**, 159–164 (1985).
7. Rea, P. A. & Poole, R. J. Proton-translocating inorganic pyrophosphatase in red beet (*Beta vulgaris* L.) tonoplast vesicles. *Plant Physiol.* **77**, 46–52 (1985).
8. Britten, C. J., Turner, J. C. & Rea, P. A. Identification and purification of substrate binding subunit of higher plant H^+ -translocating inorganic pyrophosphatase. *FEBS Lett.* **256**, 200–206 (1989).
9. Maeshima, M. & Yoshida, S. Purification and properties of vacuolar membrane proton-translocating inorganic pyrophosphatase from mung bean. *J. Biol. Chem.* **264**, 20068–20073 (1989).
10. Sarafian, V., Kim, Y., Poole, R. J. & Rea, P. A. Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized vacuolar membrane proton pump of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **89**, 1775–1779 (1992).
11. Kim, E. J., Zhen, R. G. & Rea, P. A. Heterologous expression of plant vacuolar pyrophosphatase in yeast demonstrates sufficiency of the substrate-binding subunit for proton transport. *Proc. Natl Acad. Sci. USA* **91**, 6128–6132 (1994).
12. Rea, P. A. *et al.* Vacuolar H^+ -translocating pyrophosphatases: a new category of ion translocase. *Trends Biochem. Sci.* **17**, 348–353 (1992).
13. Belogurov, G. A. & Lahti, R. A lysine substitute for K^+ . A460K mutation eliminates K^+ dependence in H^+ -pyrophosphatase of *Carboxydotherrus hydrogenoformans*. *J. Biol. Chem.* **277**, 49651–49654 (2002).
14. Li, J. *et al.* *Arabidopsis* H^+ -PPase AVP1 regulates auxin-mediated organ development. *Science* **310**, 121–125 (2005).
15. Ferjani, A. *et al.* Keep an eye on PP_i: the vacuolar-type H^+ -pyrophosphatase regulates postgerminative development in *Arabidopsis*. *Plant Cell* **23**, 2895–2908 (2011).
16. Park, S. *et al.* Up-regulation of a H^+ -pyrophosphatase (H^+ -PPase) as a strategy to engineer drought-resistant crop plants. *Proc. Natl Acad. Sci. USA* **102**, 18830–18835 (2005).
17. Guo, S. *et al.* Molecular cloning and characterization of a vacuolar H^+ -pyrophosphatase gene, SsVP, from the halophyte *Suaeda salsa* and its overexpression increases salt and drought tolerance of *Arabidopsis*. *Plant Mol. Biol.* **60**, 41–50 (2006).
18. Nakanishi, Y., Saijo, T., Wada, Y. & Maeshima, M. Mutagenic analysis of functional residues in putative substrate-binding site and acidic domains of vacuolar H^+ -pyrophosphatase. *J. Biol. Chem.* **276**, 7654–7660 (2001).
19. Hirono, M., Nakanishi, Y. & Maeshima, M. Essential amino acid residues in the central transmembrane domains and loops for energy coupling of *Streptomyces coelicolor* A3(2) H^+ -pyrophosphatase. *Biochim. Biophys. Acta* **1767**, 930–939 (2007).
20. Lee, C. H. *et al.* Identification of essential lysines involved in substrate binding of vacuolar H^+ -pyrophosphatase. *J. Biol. Chem.* **286**, 11970–11976 (2011).
21. Gordon-Weeks, R., Steele, S. H. & Leigh, R. A. The role of magnesium, pyrophosphate, and their complexes as substrates and activators of the vacuolar H^+ -pumping inorganic pyrophosphatase: studies using ligand protection from covalent inhibitors. *Plant Physiol.* **111**, 195–202 (1996).
22. Samyagina, V. R. *et al.* Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies. *J. Mol. Biol.* **366**, 1305–1317 (2007).
23. Tzeng, C. M. *et al.* Subunit structure of vacuolar proton-pyrophosphatase as determined by radiation inactivation. *Biochem. J.* **316**, 143–147 (1996).
24. Zhen, R. G., Kim, E. J. & Rea, P. A. Acidic residues necessary for pyrophosphate-energized pumping and inhibition of the vacuolar H^+ -pyrophosphatase by *N,N'*-dicyclohexylcarbodiimide. *J. Biol. Chem.* **272**, 22340–22348 (1997).
25. Pan, Y. J. *et al.* The transmembrane domain 6 of vacuolar H^+ -pyrophosphatase mediates protein targeting and proton transport. *Biochim. Biophys. Acta* **1807**, 59–67 (2011).
26. Luecke, H., Richter, H. T. & Lanyi, J. K. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* **280**, 1934–1937 (1998).
27. Van, R. C. *et al.* Role of transmembrane segment 5 of the plant vacuolar H^+ -pyrophosphatase. *Biochim. Biophys. Acta* **1709**, 84–94 (2005).
28. Buch-Pedersen, M. J., Pedersen, B. P., Veierskov, B., Nissen, P. & Palmgren, M. G. Protons and how they are transported by proton pumps. *Pflügers Arch.* **457**, 573–579 (2009).
29. Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. Crystal structure of the plasma membrane proton pump. *Nature* **450**, 1111–1114 (2007).
30. Huang, Y. T. *et al.* Distance variations between active sites of H^+ -pyrophosphatase determined by fluorescence resonance energy transfer. *J. Biol. Chem.* **285**, 23655–23664 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. F. Tam, P. C. Huang and H. J. Kung for their critical reading of the manuscript and for useful comments. The X-ray diffraction data were collected from the in-house X-ray facility at National Tsing Hua University and from beamlines BL13B1/BL13C1 at the National Synchrotron Radiation Research Center, Taiwan, and BL44XU/BL12B2 at SPring-8, Japan. This work was supported by grants from the National Science Council of Taiwan (NSC 99-2311-B-007-007-MY3 to Y.-J.S.; NSC 100-2311-B-007-001-MY3 and NSC 100-2627-M-007-012 to R.-L.P.) and National Tsing Hua University, Taiwan (99N82416E1 to Y.-J.S.).

Author Contributions S.-M.L. isolated the VrH^+ -PPase, grew the crystals and determined the structure. J.-Y.T. assisted with structural determination and completed the structural refinement. C.-D.H. assisted with structural phase analysis. C.-L.C. performed the data collection and data processing. J.-Y.T. and M.-H.L. assisted with the data collection. Y.-T.H. and T.-H.L. assisted in protein isolation. All authors participated in discussions of the results and in preparing the manuscript. Y.-J.S. and R.-L.P. supervised the project and wrote the manuscript.

Author Information The coordinates and structure factors of VrH^+ -PPase are deposited in the Protein Data Bank under the accession code 4A01. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.-J.S. (yjsun@life.nthu.edu.tw) or R.-L.P. (rlpan@life.nthu.edu.tw).

METHODS

Cloning, expression and purification. *Vigna radiata* H⁺-translocating pyrophosphatase (VrH⁺-PPase) fused with a C-terminal His-tag was expressed and isolated from *Saccharomyces cerevisiae* strain BJ2168 in accordance with a previously described method, with minor modifications³¹. Mutant derivatives were generated from wild-type VrH⁺-PPase by a QuikChange site-directed mutagenesis method³², and the sequences were verified by DNA sequencing. The VrH⁺-PPase was expressed in the yeast cells, which were transformed with the construct pYVH6 containing VrH⁺-PPase cDNA. The yeast cells were harvested by centrifugation at 4,000g for 10 min after induction for 3 days. The cell lysates were centrifuged at 4,000g for 10 min, and the supernatant was ultracentrifuged at 84,000g for 1 h for the collection of microsomal membranes. The membranes were suspended in a solution containing 25 mM MOPS-KOH pH 7.0, 20% (w/v) glycerol, 4 mM MgCl₂, 400 mM KCl and 1 mM phenylmethylsulphonyl fluoride (PMSF). The membrane proteins were solubilized from microsomal membranes with the use of n-dodecyl β-D-maltoside at a detergent-to-protein ratio of 3:1 (w/w) for 1 h at 4 °C. Insolubilized debris was removed by ultracentrifugation at 84,000g for 1 h. The solubilized VrH⁺-PPase was purified with Ni²⁺-nitrilotriacetate resin (Qiagen, Valencia) and eluted with a solution containing 25 mM MOPS-KOH pH 7.0, 20% (w/v) glycerol, 4 mM MgCl₂, 400 mM KCl, 0.15% (w/v) n-decyl β-D-maltoside (DM) and 250 mM imidazole. The purified VrH⁺-PPase was dialysed against 25 mM MOPS-KOH pH 7.0, 20% (w/v) glycerol, 4 mM MgCl₂, 0.15% (w/v) DM, 5 mM IDP, and then concentrated to 10 mg ml⁻¹ for the crystallization setup. Enzymatic activities of microsomal and purified VrH⁺-PPases were assayed in accordance with previous methods³¹.

Trypsinolysis analysis. Trypsinolysis analysis was performed as described in a previous study²⁰. The microsomal (30 μg) and purified (5 μg) proteins were incubated with L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin at a ratio of 30:1 (w/w) at 37 °C for 10 min. The proteolysis was stopped by the addition of 2% (w/w) SDS and 5 mM PMSF. The samples were subjected to western blotting analysis.

Crystallization and X-ray data collection. The hanging-drop vapour-diffusion method³³ was used to set up crystallization trials. Both VrH⁺-PPase (0.5 μl) and reservoir (0.5 μl) solutions were mixed and equilibrated against a reservoir solution (500 μl) in Linbro plates. VrH⁺-PPase crystals appeared in 2 days in the reservoir containing 23% (w/v) PEG2000, 50 mM MES pH 6.5, 10% (w/v) glycerol and 250 mM MgCl₂ at 20 °C. The native crystal data were collected on the BL44XU beamline at the SPring-8 synchrotron radiation facility, Japan. The data were processed using HKL2000 (ref. 34). VrH⁺-PPase crystals belong to the monoclinic space group C2 with the unit-cell parameters $a = 220.8 \text{ \AA}$, $b = 88.8 \text{ \AA}$, $c = 160.2 \text{ \AA}$ and $\beta = 125.9^\circ$. The Matthews's coefficient³⁵ was calculated to be $3.97 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 69% with two subunits per asymmetric unit. For phase determination, three VrH⁺-PPase heavy-atom derivatives were obtained from Ta₆Br₁₂ (2 mM), OsO₄ (1 mM) and orange-Pt (1 mM). The anomalous data

were collected on the BL13B1 and BL13C1 beamlines at the National Synchrotron Radiation Research Center, Taiwan. The data statistics are summarized in Supplementary Table 1.

Structural determination and refinement. The structural phase of VrH⁺-PPase was obtained with a tantalum derivative by MAD and from OsO₄ and Pt-organe derivatives by MIRAS (Supplementary Table 1). The heavy-atom sites were located with SOLVE³⁶. The final phase combination was calculated by AutoSol using PHENIX^{37,38}, resulting in distinguishable protein and solvent regions. The preliminary structural model was automatically built by AutoBuild^{37,38}, and the entire model was completed manually with COOT³⁹. The refinement of the structure was performed with REFMAC5 (ref. 40). PROCHECK⁴¹ was used to evaluate the stereochemistry and to assign the secondary structural elements. The structural model was refined to an *R*-factor of 16.80% and an *R*_{free} of 20.31% at 2.35 Å. The refinement statistics are summarized in Supplementary Table 1. All the figures were generated using PYMOL⁴². The solvent-accessible surface area was calculated with the program HOLLOW⁴³ using a 1.4-Å probe radius.

- Hsu, S. H. *et al.* Purification, characterization, and spectral analyses of histidine-tagged vacuolar H⁺-pyrophosphatase expressed in yeast. *Bot. Stud. (Taipei, Taiwan)* **50**, 291–301 (2009).
- Kirsch, R. D. & Joly, E. An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.* **26**, 1848–1850 (1998).
- McPherson, A. Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189**, 1–23 (1990).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
- DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific) (2002).
- Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature11029

Corrigendum: Recovery rates reflect distance to a tipping point in a living system

Annelies J. Veraart, Elisabeth J. Faassen, Vasilis Dakos, Egbert H. van Nes, Miquel Lüring & Marten Scheffer

Nature **481**, 357–359 (2012)

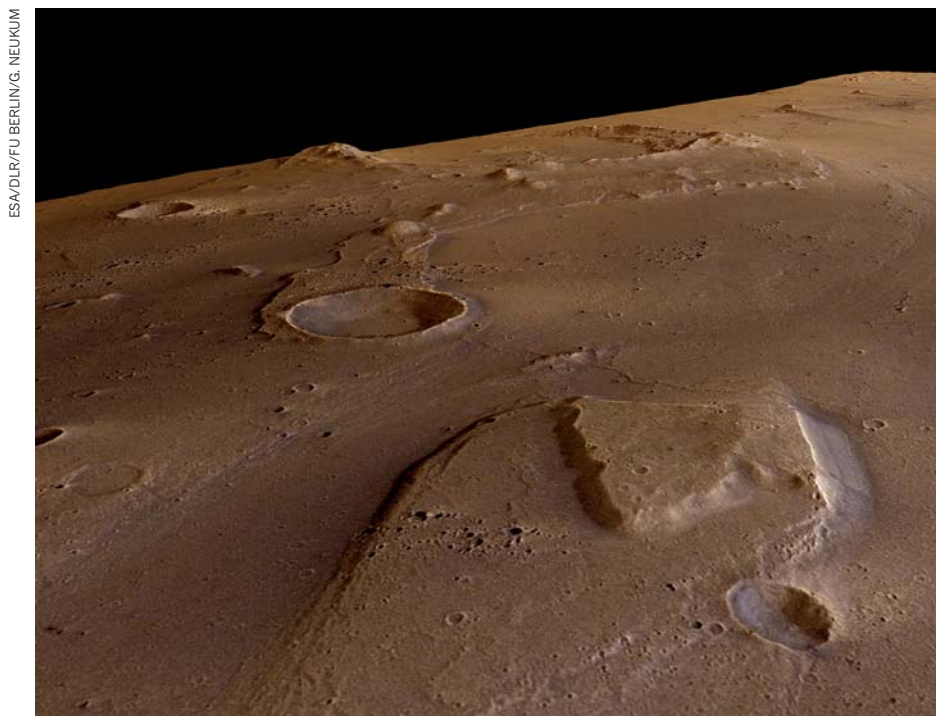
There was a scaling error in the light attenuation values of Fig. 1 and Supplementary Fig. 2.1 (about 210 should have been subtracted from each value). Also, the sentence “We perturbed the populations every 4–5 days by removing 10% of their biomass through dilution.” on page 357 should have read “We perturbed the populations every 4–5 days by flushing with medium (10% of the volume), which was equivalent to a reduction of the biomass by 3–5% owing to incomplete mixing.”. This sentence and Fig. 1 and Supplementary Fig. 2.1 have been replaced in the PDF and HTML versions online. These changes do not alter any of the conclusions of this Letter. We thank J. Huisman for drawing our attention to these issues.

CAREERS

TURNING POINT Postdoc's software code for imaging pays career dividends **p.407**

INTERVIEWS Narcissists are more likely to excel in job interviews, says study **p.407**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



Images from the surface of Mars: analysing these may form part of an astrobiologist's role.

ASTROBIOLOGY

Alien encounters

Establishing a career in astrobiology is not always easy. But keeping one's options open can lead to work in this niche field.

BY NAOMI LUBICK

Michael Callahan knows that he has a great job every time a meteorite sample arrives at the laboratory from NASA's meteorite archive in Houston, Texas. "There's no cooler feeling than working with a meteorite in the lab," he says.

Callahan, an analytical chemist at the Astrobiology Analytical Laboratory at NASA's Goddard Space Flight Center in Greenbelt, Maryland, is one of a growing number of researchers around the world working in astrobiology — the exploration of how and where life might have emerged in the Universe.

The breadth of astrobiology can be daunting. Because the field focuses on questions about the origins of life, it incorporates

aspects of astronomy and physics as well as chemistry, geology, oceanography, microbiology and bioinformatics. As a result, there are many different routes to a career in astrobiology; Callahan, for example, started out as an analytical chemist in industry. Exobiology — the study of the possibility of life on other planets — thrived with origins-of-life research in the 1950s and '60s, and really took off with the space exploration programmes in the 1960s and '70s. Later, when the astrobiology label took hold, some people questioned the field's seriousness, notes Zita Martins, an astrobiologist at Imperial College London. Many different types of researchers now contribute to the field; geomicrobiologists look for extremophiles — organisms that live in extreme environments such as deep caves or hot

springs; astronomers search for chemical signs of metabolism on other planets; and biologists investigate how life on Earth first appeared.

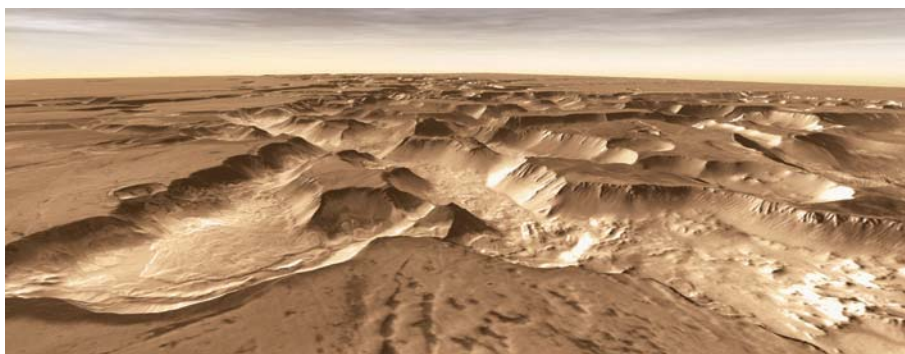
Martins, for example, works in the department of Earth science and engineering, developing techniques for uncovering microbial life and metabolism in extreme settings, using soil samples from the Utah desert as stand-ins for samples from Mars. She collaborates with soil scientists, geochemists, biologists and robotics engineers who specialize in exploration. "That's the beauty of the astrobiology field. It's scientists from different research fields, working together," she says. NASA has an annual conference dedicated to the field, which is taking place this week in Atlanta, Georgia.

JOINT MISSIONS

Support for astrobiology research is spread across many funding streams and disciplines, and finding funding and jobs that are strictly astrobiology-related is difficult, especially given recent budget woes. NASA and the European Space Agency (ESA) support the largest number of astrobiology-related positions, whether they are in Earth-science observation or engineering jobs on the next space mission. But NASA is facing budget cuts that have alarmed researchers in the field. For 2012 and 2013, a slight increase in the budget for human space exploration, from US\$3.7 billion to a proposed \$3.9 billion, was countered by a decrease for the science directorate, including planetary-science research, from \$5.1 billion to a proposed \$4.9 billion. The agency has withdrawn its component of the ExoMars programme — a joint mission with ESA to send an orbiter and rover to look for methane on the red planet, which is scheduled for 2018. NASA's funding would have created more research positions for biologists and instrumentalists, among others.

Although the US Terrestrial Planet Finder — a mission intended to look for Earth-sized planets in the habitable zones of distant solar systems — has been suspended, astronomers hope that NASA's James Webb Space Telescope, an observational mission with similar planet-finding goals, will continue moving forward.

Despite cuts in some areas, NASA's Earth budget is up slightly from \$1.76 billion in 2012 to \$1.79 billion in 2013, notes geochemist Mitchell Schulte, a programme officer for the Mars Exploration Programme at NASA headquarters in Washington DC. That means that research done on this planet but with applications off-planet could still be awarded funding, such as hunting for extremophiles and ►



Noctis Labyrinthus, a region of the red planet captured by NASA's Mars Odyssey orbiter.

their proteins, or remote-sensing research such as LiDAR (a technique for detecting water, or mapping minerals and topography).

The Centre for Astrobiology in Madrid, which was established in 2000 as a partner to the NASA Astrobiology Institute based in Mountain View, California, is an example of an organization that is crucial for strengthening ties between researchers around the world, says astrophysicist Lisa Kaltenegger, who holds dual positions at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and the Max Planck Institute for Astronomy in Heidelberg, Germany.

"I look for people who are willing to take this risk. In my experience, that is an extraordinarily hard thing to do."

In addition to NASA's partner centre in Spain, other centres have sprung up, including Germany's Institute of Planetary Research in Berlin, which is a cooperative alliance of about 200 scientists across 17 science organizations, with funding from the Helmholtz Alliance based in Bonn, which studies planetary evolution. Gerda Horneck, a former deputy director of the German Aerospace Centre's Institute of Aerospace Medicine in Cologne, says that there are now 25 PhD candidates in the Helmholtz Space Life Sciences Research School, based in various locations in Germany.

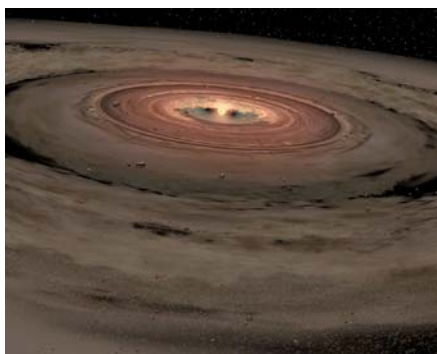
For fledgling scientists in the field, the economic climate means that success can hinge on finding multiple funding sources. Julie Huber, a microbial oceanographer at the Marine Biological Laboratory in Woods Hole, Massachusetts, whose research focuses on life at deep hydrothermal vents, says she applies for grants from NASA and the US National Science Foundation for oceanography programmes and occasionally from the US National Institutes of Health for microbiology research. She also applies to private foundations such as the Gordon and Betty Moore Foundation in California and the Alfred P. Sloan Foundation in New York. Huber says she considers herself lucky to have secured funding, given the budget woes of government agencies and the state of the foundations' endowments. "It used to be you had to write one or two grants to support yourself for five years,"

she says. "Now it's four, five, six grants every year, with no growth in the pool of money."

Other initiatives have helped to augment support for the field. The CAREX project from the European Union's Seventh Framework programme for research, which coordinated the study of life in extreme environments, wrapped up this year after spending €1.4 million (US\$1.8 million). Its data on desert life, specifically microbial diversity and aspects of life in extreme environments, could help to inform surveys for life on Mars, although it is not labelled as astrobiology. The next round of European funding, Horizon 2020, will include support especially for astrobiology and space exploration, says Horneck. How much support is still under discussion. Up to this point, various projects across partnered universities and institutes have received about €2 million over three years from the European Union for cooperative astrobiology research, the CAREX funding being one example.

SPACE SCHOOL

Few PhD programmes award astrobiology degrees; students typically earn their degree in a field such as analytical chemistry or microbiology but focus on astrobiology for their thesis. For example, the University of Washington in Seattle confers a certificate in astrobiology for graduate-level research, but graduate students retain their home-department affiliation for their doctorate. Graduates with this certificate have gone on to work as astrobiologists or in related fields at NASA or other institutes.



Astrobiologists hope to continue to study stars such as 51 Ophiuchi (artist's rendition pictured).

In 2005, Stockholm University launched an astrobiology graduate school for an initial five years to gauge interest in the field, a sort of interdisciplinary curriculum experiment. The university considered it a success, and departments including astronomy and geology have transferred their experience into the newly created astrobiology centre.

Sandra Siljeström's career exemplifies the potential myriad roles an astrobiologist can have. A graduate from the Stockholm programme who trained as a geochemist, Siljeström is now a research scientist at the government-run SP Technical Research Institute of Sweden in Borås, where she works on projects for the Swedish space board, developing methods to examine small amounts of organic compounds using time-of-flight spectroscopy. She is also helping to develop the Mars Organic Molecule Analyser, an instrument to be carried on the ExoMars rover. Her position also includes contract work for industry, examining organic compounds in rocks and other materials.

MORE BODIES

Horneck advises undergraduates and master's students to ground themselves in one of the many applicable sciences before moving on to astrobiology-related research. She says getting a solid grounding in their field before pursuing astrobiology is important.

However, Malcolm Walter, director of the Australian Centre for Astrobiology at the University of New South Wales, advises students to start taking astrobiology courses at the undergraduate level if they are available. "The skills are generic and can be applied to many career paths. Being interdisciplinary, thinking in an interdisciplinary fashion leads to many career opportunities," he says. "I look for people who are willing to take this risk. In my experience, that's an extraordinarily hard thing to do."

Walter says two of his PhD students recently ended up working for NASA — one as a classical geologist at the Jet Propulsion Laboratory in Pasadena, California, and the other as a mineral spectroscopist, mapping mineralogy on the surface of Mars. Other students have gone on to work for biotechnology firms or the booming Australian mining industry.

Horneck worries that talented graduates will leave the field because few jobs exist outside NASA, ESA and academia, and even there, positions are limited. Many projects are suited to PhD theses, but not to full-time jobs. Not enough opportunities exist for bright doctorates to continue their research. "This is something that needs attention," she says. Even so, student interest portends good things for the field. "Ten years ago, there were no courses. Now it's a field very much taken up by students," Horneck says. Despite the challenges, she foresees a burgeoning field, with more discoveries on the horizon. ■

Naomi Lubick is a freelance writer based in Stockholm, Sweden.

TURNING POINT

Marc Modat

A postdoc in non-linear medical-image registration at University College London, Marc Modat has created imaging software that has helped him to co-author more than 60 publications. In February, the software landed him the university's Young Investigator in Neuroimaging Techniques Award for 2011.

You hadn't planned to do a PhD. How did you end up doing one?

Towards the end of my undergraduate degree in bioengineering at the University of Creteil in Paris, I did a medical-imaging internship at the Commonwealth Scientific and Industrial Research Organisation near Brisbane, Australia, with Sebastien Ourselin, who was working on how to compare the brain scans of patients. I had never considered doing a PhD, but he was moving to University College London and he asked if I wanted to join his group. This turned out to be one of the best career decisions I've made.

Why did you pursue neuroimaging?

I didn't have strong feelings about medical imaging, I just met good people at a good time, and I enjoyed working on the algorithms needed to align and compare the scanned images of patients. This is called medical-image registration and is a crucial piece of the image-processing pipeline. For example, if you want to compare the brain scans of two different patients or the scans of the same patient over time, you need to have a way to quantify any changes.

What was so popular about the software code you designed?

Computation time was a big challenge in image processing when I started. We were working with large cohorts of patients, and it could take hours to process images. I found a way to make the technique up to 200 times faster but without compromising accuracy, and I put the code online for other scientists. I'm not saying that my code is better than the others available online, but it may be faster, which got people's attention. I received a great deal of peer recognition, invitations to be part of collaborations and the young investigator award because of the code.

Were there objections to publishing the code?

Not really. Not everyone's supervisor would let you do that — some would prefer to maintain their competitive advantage — but we decided to publish the code because we knew



that it would increase our visibility. It did. I receive e-mails from people all over the world saying that they are using it or suggesting modifications. We are still working on it daily and adding more features to ensure accuracy and increase its speed.

Are all your publications a result of developing the code?

Most are. As a PhD student, I co-authored publications with other research teams because our code and imaging expertise were used to help with their work. I am part of a large laboratory, so I have been involved in a lot of different clinical studies, but now I'm trying to find a balance between working on my projects and working as a collaborator.

Do you think your career interests will dictate the development of the code?

Going forward, I'm not sure the code will be linked only to me; many people are contributing to it by suggesting changes or additional features. I would like to be involved in everything related to the code, but that depends on how it grows. We are thinking of merging the medical-image registration software, the software that reconstructs images, and the simulation programs and visualization tools into one package, which could have a big effect on the field.

Why is this award important for you?

Supervisors nominate their students every year for this award. Typically, it goes to a clinician or a biologist. When I received the award, it made me think that my work is as valued as that of the clinicians — which is nice because I really enjoy what I do. ■

INTERVIEW BY VIRGINIA GEWIN

INTERVIEWS

Narcissists excel

Job hunters who hope to do well during interviews should talk about themselves in glowing terms, says a study in the press in the *Journal of Applied Social Psychology*. The authors found that behaviours that would be considered narcissistic in other settings were rated highly in simulated job interviews. "This is one setting where humility is not valuable," says study co-author Peter Harms, assistant professor of management at the University of Nebraska-Lincoln. He encourages interviewees to speak a lot. "Don't give short answers," he says. Those who spoke quickly, smiled, gestured and complimented others often, and who spent time on grooming were rated highly.

FUNDING

Global-health research

A fellowship programme will award US\$20.3 million over 5 years to 400 early-career scientists in the United States and in low- and middle-income countries to build the next generation of global-health researchers. Eighteen centres at the US National Institutes of Health (NIH) in Bethesda, Maryland, including the Fogarty International Center are supporting the fellowships. Participants will be selected by their institution to conduct research at one of about 80 sites globally. Trainees will study issues such as HIV/AIDS, tuberculosis, malaria, maternal and child health, cancer, cardiovascular disease and diabetes. Postdocs will make up 80% of new fellows and graduate students the remaining 20%.

INVESTMENT

UK drug development

A £250-million (US\$397-million) investment in venture capital for translational research in the United Kingdom could create jobs. The European Investment Fund in Luxembourg City and Cancer Research Technology (CRT), the commercial arm of the London-based charity Cancer Research UK, are jointly investing £50 million in the development of potential cancer drugs. A separate £200-million initiative by the Wellcome Trust in London will support early-stage biotechnology firms. Keith Blundy, chief executive of CRT, says that experts in drug development should find opportunities as the funding helps to take more drug candidates through to clinical trials.

MONKEYS

Prose by any other name.

BY KEN LIU

Ted and Kathy stared at the chaotic scene through the bars of the cage. A large, male macaque monkey about two feet tall screeched and lifted the typewriter — a 1953, lime green Olivetti Lettera 22 — over his head. He stood still for a second like a weight-lifter, and then threw the machine hard against the ground. It clattered on the dirt floor, keys and platen jangling, and came to rest with the sequence “jl,dykb nvxcliuear cx,juoun2 ep89 xadl” on the sheet of paper.

Kathy's hands covered her mouth. She blew out a breath. “At least we now have our first keystrokes.”

Ted just shook his head.

Two smaller males approached the typewriter. One jumped up and down on the keys: “cx,juoun2 ep89 xadl.” The other watched, and then decided to defecate into the curved bowl formed by the type bars.

“Well, Professor Emroche isn't going to want *that* typewriter back,” Kathy said, after recovering from the initial shock. Then she started to laugh. The monkeys stopped to stare at her, which only made her laugh harder.

Ted shook his head again.

The ‘Monkey Shakespeare’ project was an interdisciplinary collaboration between the humanities and the science departments. But after a short clip of the monkeys pooping into the typewriter was circulated online, everyone began to distance themselves from it.

“We don't really think there's much *science* in a project like this,” said Professor Kun of the Department of Computer Science. “That old chestnut about an infinite number of monkeys banging on an infinite number of typewriters is just a thought experiment. You should really ask the literature people what they're trying to accomplish with this.”

“This is obviously not how we envisage the future study of *literature*,” said Professor Emroche of the Department of American and English Language and Literature. “We already have the works of Shakespeare, so why would we want monkeys reproducing it by chance? You should really talk to the biology people about what they are trying to get out of it.”

In the end, Ted and Kathy, undergraduate assistants who had yet to declare their majors



and who had signed on to feed the monkeys, became the only names attached to the project. The students felt like orphans. No one wanted to claim responsibility for them.

They decided to treat the project as a piece of performance art.

After a while, the monkeys stopped abusing the typewriter. They mainly left it alone during the day, like a toy they found boring. But once in a while, one or a group of them would come by and bang on the keys until they grew tired of it again. Ted and Kathy stopped by several times a day to swap out filled pages for fresh ones.

Kathy flipped through the stack of typescript they had collected, examining each like a puzzle.

“There's a word!” she said.

Ted looked at where she was pointing. Amid a full page of jumbled letters, a sequence of four stood out like a bright diamond: “734q9u8opfuiou wise djk;we897d78.”

Ted sighed. “It took the monkeys 5 days to produce 25 pages, and we have one word. At this rate, we aren't going to have much to show by the end of the semester. Not much of an art project.”

“You're missing the point about art,” Kathy

said. She threw the papers up in the air and watched them drift to the ground like a flock of pigeons. “It's not about what you have to show, but how you interpret what you have.”

“How would you interpret this?”

“It shows just how rare order is in this random universe we live in. It shows how art is precious and miraculous. It shows the true measure of Shakespeare's genius.”

Ted laughed. “Yeah, that B. S. sounds pretty good.”

“It's the viewer who gives meaning to art,” Kathy said, laughing as well. “Just like it's the reader who gives meaning to the book.”

Wflq rushed into the Eqqlb, gesticulating wildly.

She waved a sheaf of papers in front of 7*uLw, who asked her calmly what in the world was the matter with her.

Wflq was too excited to speak coherently. It took a while before 7*uLw finally understood that her experiment had finally yielded some results. In one of the infinite number of universes that she had been observing, the native creatures had finally produced something worthy of notice.

Although the natives who dominated that world seemed to bang on keyed machines randomly all the time, this time they had allowed some other creatures — deemed inferior by them for some reason — to use one of their machines. These ‘lesser’ creatures immediately produced the most beautiful poetry that the universe had ever created. Out of the babbling, random chaos that made up almost all the symbolic output of that world, this new string stood out like a clear voice in the wilderness.

But tragically, the natives of that universe did not seem to understand what they had. Wflq had to rescue a great piece of art.

Here's how the masterpiece began: “jl,dykb nvxcliuear cx,juoun2 ep89 xadl...”

As they read, Wflq and 7*uLw fell in rapture. The beauty of the language overwhelmed them.

Translated into our language, the book Wflq produced started this way:

Boatswain!

Here, master: what cheer? ■

Ken Liu lives and writes in Massachusetts. You can read more of his fiction at <http://kenliu.name/stories>.

JACEY

A *Xanthomonas* uridine 5'-monophosphate transferase inhibits plant immune kinases

Feng Feng^{1,2,3}, Fan Yang³, Wei Rong⁴, Xiaogang Wu³, Jie Zhang³, She Chen³, Chaozu He^{2,4} & Jian-Min Zhou^{3,5}

Plant innate immunity is activated on the detection of pathogen-associated molecular patterns (PAMPs) at the cell surface, or of pathogen effector proteins inside the plant cell^{1–4}. Together, PAMP-triggered immunity and effector-triggered immunity constitute powerful defences against various phytopathogens. Pathogenic bacteria inject a variety of effector proteins into the host cell to assist infection or propagation. A number of effector proteins have been shown to inhibit plant immunity⁵, but the biochemical basis remains unknown for the vast majority of these effectors. Here we show that the *Xanthomonas campestris* pathovar *campestris* type III effector AvrAC enhances virulence and inhibits plant immunity by specifically targeting *Arabidopsis* BIK1 and RIPK, two receptor-like cytoplasmic kinases known to mediate immune signalling^{6–8}. AvrAC is a uridylyl transferase that adds uridine 5'-monophosphate to and conceals conserved phosphorylation sites in the activation loop of BIK1 and RIPK, reducing their kinase activity and consequently inhibiting downstream signalling.

Xanthomonas campestris pathovar *campestris* (Xcc) is a causal agent of black rot diseases on numerous crucifer plants such as *Brassica* and *Arabidopsis*. AvrAC exists in all three sequenced strains of Xcc and a newly sequenced *Xanthomonas campestris* pathovar *raphani* strain, but not other *Xanthomonas* species. (<http://www.xanthomonas.org/t3e.html>). In Xcc strain 8004 (Xcc8004), AvrAC is recognized by *Arabidopsis* plants to trigger a weak effector-triggered immunity (ETI) in vascular, but not mesophyll tissues⁹. To determine if AvrAC has a role in virulence, we generated an *avrAC* knockout strain (Xcc8004 Δ avrAC) and performed competitive index assays on cabbage plants (Methods). The competitive index values (Xcc8004 Δ avrAC to Xcc8004 ratios) were 0.5–0.3 in 3/5 cultivars tested in multiple experiments (Fig. 1a), demonstrating a role of AvrAC in virulence on these plants. Well known PAMPs, including the bacterial flagellar peptide flg22, bacterial elongation factor (EF-Tu) peptide elf18, and fungal cell wall component chitin¹⁰, are recognized by pattern recognition receptors (PRRs) FLS2, EFR and CERK1, respectively⁴, to induce the expression of a PAMP-responsive reporter gene *FRK1::LUC*. The presence of the AvrAC transgene largely inhibited *FRK1::LUC* induction in protoplasts (Fig. 1b and Supplementary Fig. 1), indicating that AvrAC is a potent inhibitor of the PAMP-triggered immunity (PTI) signalling pathway.

Two AvrAC transgenic lines were generated to test if AvrAC impedes mesophyll resistance to Xcc8004, a nonpathogenic mutant bacteria Xcc8004 Δ hrcV and *Pseudomonas syringae* pathovar *tomato* (*Pst*) Δ hrcC mutant bacteria. In all three cases, the transgenic lines supported 0.7–1.1 log more bacterial growth than did the wild-type plants (Fig. 1c, d). Thus AvrAC can markedly compromise PTI resistance to these bacteria. The AvrAC transgenic lines were also inoculated with *Pst* strains carrying effector genes *avrB*, *avrRpt2* or *avrPphB*, which elicit ETI resistance specified by the cytoplasmic immune receptors RPM1, RPS2 and RPS5, respectively³. The transgenic lines supported 0.9–1.0 log more *Pst*(*avrB*) bacterial growth than did wild-type

plants (Fig. 1e). In contrast, the transgenic lines supported less than 0.3 log increase in bacterial growth when *Pst*(*avrRpt2*), *Pst*(*avrPphB*) and *Pst* bacteria lacking an *avr* gene were used (Fig. 1e and Supplementary Fig. 2). Thus, although the AvrAC transgene compromised RPM1-dependent ETI, it did not appear to affect RPS2- and RPS5-dependent ETI. To determine further if bacterially delivered AvrAC similarly interferes with RPM1 activation, we infiltrated *Arabidopsis* plants with Xcc8004 before *Pst*(*avrB*) inoculation. Whereas the control plants inoculated with *Pst*(*avrB*) alone developed strong a hypersensitive

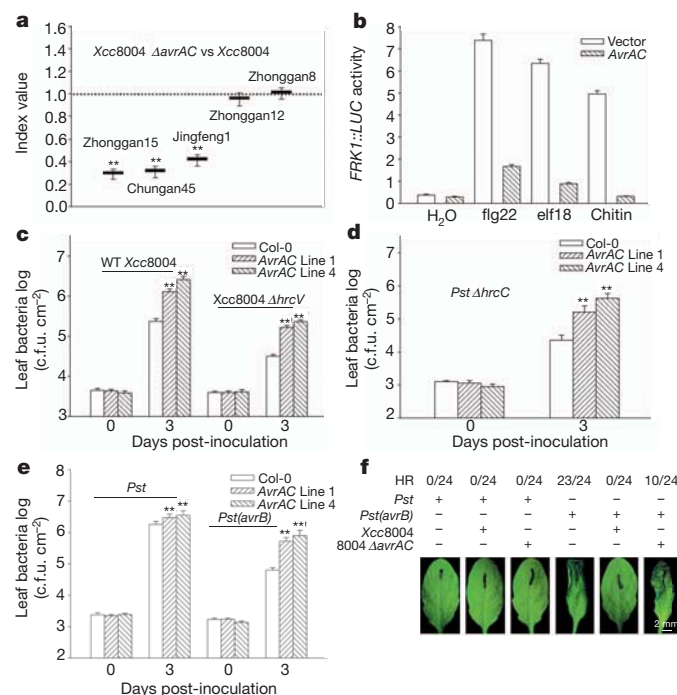


Figure 1 | AvrAC inhibits plant immunity. **a**, *avrAC* promotes Xcc bacterial virulence on cabbage plants. Competitive index value represents an average ratio of Xcc8004 Δ avrAC to Xcc8004 bacterial growth on the indicated cabbage cultivars (mean \pm s.d.; $n \geq 6$; ** $P < 0.01$, Student's *t*-test). **b**, Transient expression of AvrAC-Flag in protoplasts inhibits PAMP-induced *FRK1::LUC* transcription (mean \pm s.d.; $n \geq 3$). **c–e**, The AvrAC transgene compromises PTI and RPM1-specified ETI in plants. Wild-type (WT; Col-0) and AvrAC transgenic lines pre-treated with oestradiol were infiltrated with Xcc8004 and Xcc8004 Δ hrcV (**c**), *Pst* Δ hrcC (**d**), or *Pst* or *Pst*(*avrB*) (**e**). Bacterial population in the leaf was measured at the indicated times (mean \pm s.d.; $n \geq 6$; ** $P < 0.01$, Student's *t*-test). **f**, Bacterially delivered AvrAC inhibits the RPM1-specified hypersensitive response (HR). Wild-type *Arabidopsis* plants were infiltrated with the indicated Xcc strains 1 day before infiltration of *Pst* or *Pst*(*avrB*), and development of the hypersensitive response was photographed 6 h later. The ratio of leaves showing a hypersensitive response to the total number of injected leaves is shown on top.

¹School of Life Sciences, Tsinghua University, Beijing 100084, China. ²Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China. ³National Institute of Biological Sciences, Beijing 102206, China. ⁴Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, Hainan University, Haikou, Hainan 570228, China. ⁵State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China.

response, the plants pre-treated with *Xcc8004* developed no hypersensitive response (Fig. 1f), indicating a strong interference of RPM1 activation by *Xcc8004*. The *Xcc8004ΔavrAC* mutant strain was less capable of preventing hypersensitive response development (Fig. 1f and Supplementary Fig. 3), indicating that AvrAC interferes with RPM1-mediated ETI when delivered from *Xcc*.

To identify host targets and the mechanism by which AvrAC inhibits PTI, we systematically investigated various molecular events of PTI signalling pathways¹¹. flg22 perception triggers H₂O₂ production and MAP kinase (MAPK) activation in wild-type plants within several minutes; both events were completely abolished or diminished in AvrAC transgenic plants (Fig. 2a, b). However, AvrAC was unable to inhibit MAPK activation triggered by the constitutively active MKK5^{DD} (ref. 10; Supplementary Fig. 4). These results suggested that AvrAC acts upstream of MAPK cascades and H₂O₂ production to inhibit PTI signalling.

PAMP detection rapidly induces phosphorylation of BIK1 (refs 6, 7), which acts immediately downstream of FLS2, EFR and CERK1 (ref. 7). The expression of AvrAC in protoplasts prevented the flg22-induced BIK1 phosphorylation, as indicated by a slow migrating form on SDS-polyacrylamide gel electrophoresis (SDS-PAGE; Fig. 2c). Instead, BIK1 uniformly migrated between the unphosphorylated and phosphorylated forms independent of flg22 treatment.

Upstream of BIK1 phosphorylation is the flg22-induced association between FLS2 and a receptor-like kinase called BAK1, forming an active receptor complex^{12,13}. This interaction occurred normally in protoplasts irrespective of the presence or absence of AvrAC (Fig. 2d).

The results described above suggested that AvrAC directly target BIK1, FLS2 or BAK1. Co-immunoprecipitation assays showed that AvrAC can interact with BIK1, but not FLS2 or BAK1 (Fig. 2e). Glutathione S-transferase (GST) pull-down assays indicated that this interaction is direct (Fig. 2f). RIPK, an RLCK related to BIK1, can phosphorylate the RPM1-interacting protein RIN4 at T166 (ref. 8), and the phosphorylated RIN4 is responsible for triggering RPM1-specified ETI^{8,14}. AvrAC also interacted with RIPK *in vitro* and in

protoplasts (Figs 2f and 4b). We conclude that BIK1 and RIPK are specific targets of AvrAC, a conclusion consistent with its inhibition of PTI and the RPM1-specific ETI.

The AvrAC-induced electrophoresis mobility shift of BIK1 was resistant to phosphatase treatment (Supplementary Fig. 5a). Co-expression of BIK1 and AvrAC in *Escherichia coli* detected the same phosphatase-resistant, slow migrating form of BIK1 (Supplementary Fig. 5b, c), suggesting an AvrAC-induced BIK1 modification that is not caused by protein phosphorylation.

AvrAC contains a filamentation-induced by c-AMP domain (Fic) in the carboxy terminus. More than 4,300 Fic-domain-containing proteins have been found in bacteria and animals, mostly with unknown biochemical function^{15,16}. Two Fic-domain-containing, animal bacterial pathogen effectors *Vibrio* VopS and *Histophilus* IbpA catalyse adenylation (AMPylation) on threonine or tyrosine residues of Rho family GTPases to disrupt cytoskeleton in the host cell^{17,18}. Another bacterial effector protein, *Legionella pneumophila* DrrA (also known as SidM) uses a glutamine synthetase adenylyl transferase domain to adenylylate another host GTPase called Rab1b¹⁹. However, *in vitro* labelling and mass spectrometric assays with AvrAC and BIK1 recombinant proteins failed to detect any AMPylation on BIK1. Surprisingly, mass spectrometry of the AvrAC recombinant protein (expressed alone in *E. coli*) revealed multiple modified protein components with serial additions of 306 Da, indicative of uridine 5'-monophosphate (UMP) modifications (Supplementary Fig. 6a). Tandem mass spectrometry (MS/MS) revealed multiple serine and threonine residues in AvrAC containing mono-UMP modifications (Supplementary Fig. 6b). The results suggested that AvrAC is a UMP transferase. *In vitro* assays with α^{32} P-UTP showed that both BIK1 and RIPK were uridylylated only when incubated with AvrAC (Fig. 3a). AvrAC also labelled itself in this assay, explaining the UMP modifications on AvrAC described above. PBL1, a homologue of BIK1 that acts additively with BIK1 in PTI signalling⁷, was also modified by AvrAC (Supplementary Fig. 7). Interestingly, another RLCK VII subfamily member, PBS1, was not modified under

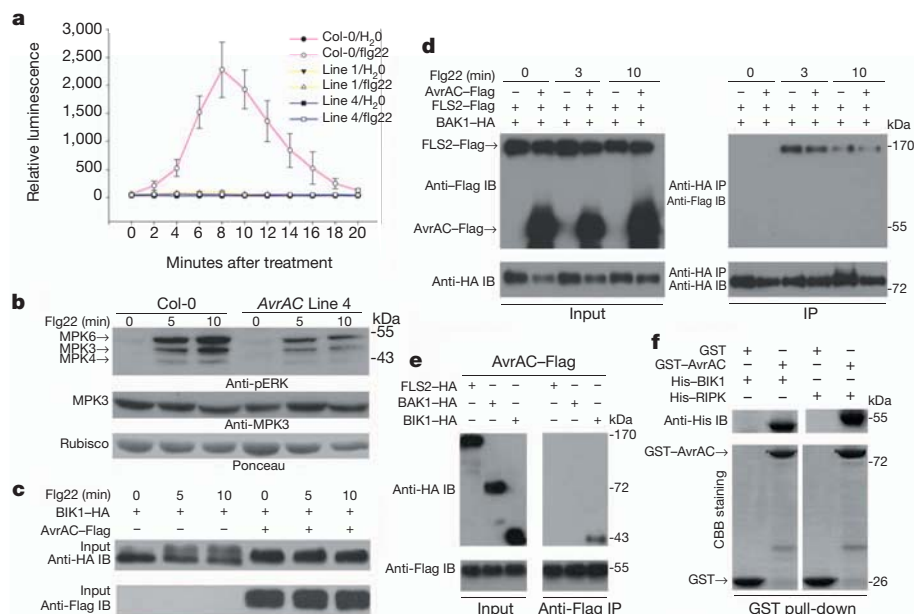


Figure 2 | AvrAC targets BIK1 and inhibits early PTI signalling events.

a, Flg22-induced H₂O₂ production in wild-type and AvrAC transgenic plants (mean \pm s.d.; $n \geq 8$). **b**, Flg22-induced MAPK phosphorylation in wild-type and AvrAC transgenic plants. The identities of phosphorylated MAPKs are labelled as described (Methods). Equal loading is indicated by immunoblot with anti-MPK3 antibodies or ponceau staining of Rubisco protein. **c**, Flg22-induced BIK1-HA phosphorylation in protoplasts in the presence or absence of AvrAC-Flag. **d**, Flg22-induced FLS2-BAK1 association in protoplasts in the

presence of AvrAC as determined by co-immunoprecipitation assays. The minor differences of BAK1-HA levels resulted from sample processing. **e**, AvrAC interacts with BIK1, but not FLS2 and BAK1, in protoplasts. BIK1-HA, BAK1-HA and FLS2-HA were co-expressed with AvrAC-Flag in protoplasts and subjected to co-immunoprecipitation. **f**, AvrAC interacts with BIK1 and RIPK *in vitro* in a GST pull-down assay. The bound proteins were analysed by anti-His immunoblot and the membrane stained with Coomassie blue (CBB). IB, immunoblot; IP, immunoprecipitation.

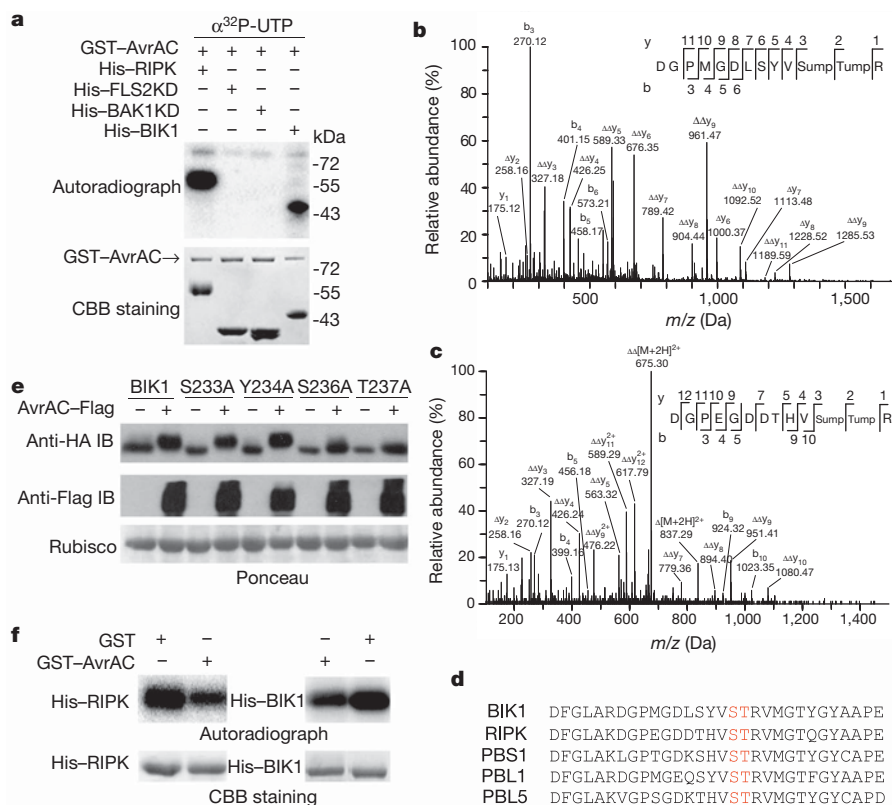


Figure 3 | AvrAC uridylylates BIK1 and RIPK. **a**, AvrAC uridylylates BIK1 and RIPK, but not FLS2 and BAK1, *in vitro*. GST-AvrAC was incubated with His-BIK1, His-RIPK, or His-tagged kinase domains (KD) of FLS2 or BAK1, in the presence of $\alpha^{32}\text{P}$ -UTP. The resulting products were subjected to SDS-PAGE and autoradiography. **b**, **c**, AvrAC uridylylates conserved serine and threonine residues in the activation loop of BIK1 and RIPK in the plant cell. BIK1-Flag (**b**) or RIPK-Flag (**c**) was co-expressed with AvrAC-Flag in protoplasts, affinity purified, and subjected to MS/MS spectrometric analyses. The b- and y- ions are marked in the spectrum and illustrated along the peptide sequence above the spectrum. This unambiguously identified BIK1 Ser 236 and Thr 237, and RIPK Ser 251 and Thr 252 as UMP-modified residues. **d**

represents neutral loss of an UMP group with nominal mass of 324 Da.

d, Alignment of the activation loop of selected RLCK VII kinases. The UMP modification sites are highlighted in red. **e**, BIK1 Ser 236 and Thr 237 are required for uridylylation by AvrAC. Various BIK1-HA mutants carrying the indicated amino acid substitutions were co-expressed with AvrAC-Flag and analysed by immunoblot. Note the increased BIK1 protein level in the presence of AvrAC, suggesting an AvrAC-induced stabilization of BIK1 protein. **f**, Modification by AvrAC inhibits BIK1 and RIPK kinase activity. His-BIK1 or His-RIPK co-expressed with GST-AvrAC or GST in *E. coli* was affinity purified and subjected to autophosphorylation assay.

the same condition. The same assay failed to label the kinase domain of FLS2 and BAK1, indicating that certain members of the RLCK VII subfamily are specific substrates of AvrAC (Fig. 3a). MS/MS analysis on the recombinant BIK1 protein that had been co-expressed with AvrAC uncovered a tryptic peptide (DGPMGDLSYVSTR) derived from the activation loop containing UMP at Ser 236 and Thr 237 (Supplementary Fig. 8a). Likewise, RIPK that had been co-expressed with AvrAC also produced a tryptic peptide (DGPEGDDTHVSTR) located in the activation loop in which Ser 251 and Thr 252 were UMP modified (Supplementary Fig. 8b). BIK1 and RIPK proteins that had been co-expressed with AvrAC in protoplasts also contained the same UMP modifications (Fig. 3b, c), indicating that AvrAC can uridylylate BIK1 and RIPK in the plant cell. The vast majority of the BIK1 and RIPK activation loop peptides were uridylylated (Supplementary Fig. 9). Together these results unequivocally demonstrated that AvrAC is a potent enzyme that uridylylates BIK1 and RIPK both *in vitro* and in plant cells.

ATP is the preferred nucleotide substrate for IbpA and VopS to modify Rho GTPases *in vitro*¹⁶, although GTP, UTP and CTP can also be used. Whether nucleotides other than ATP are used by these effectors to modify Rho GTPases in host cells remains unknown. AvrAC failed to AMPylate BIK1, and we observed no AMPylation on any peptides of BIK1, RIPK or AvrAC expressed in *E. coli* or plant cells. Another Fic-domain protein, AnkX, from *Legionella pneumophila* was recently shown to covalently attach a phosphocholine moiety to the Rho GTPase Rab1 (ref. 20). The commonality shared by the three

modifications suggests that Fic-domain proteins are phosphotransferases with different substrates.

BIK1 Ser 236 and Thr 237 and RIPK Ser 251 and Thr 252 are highly conserved among RLCK VII subfamily members (Fig. 3d). Phosphorylation of BIK1 Ser 236 and Thr 237 has a crucial role in PTI signalling^{6,7,21}. AvrAC failed to induce the BIK1(S236A) and BIK1(T237A) mutant protein migration in protoplasts (Fig. 3e), indicating that these residues are required for UMP modification. A similar result was obtained with a BIK1(S233A,Y234A,S236A,T237A) quadruple mutant⁷ (Supplementary Fig. 10). In contrast, mutations in Lys 105, an invariant residue required for ATP binding, and Ser 233 and Tyr 234, two less conserved amino acid residues in the activation loop, did not affect the AvrAC-induced migration (Supplementary Fig. 10 and Fig. 3e). RIPK and BIK1 that had been UMP modified by AvrAC in *E. coli* displayed reduced autophosphorylation (Fig. 3f). Taken together, these results demonstrated that AvrAC inhibits BIK1 and RIPK kinase activities by directly uridylylating the conserved phosphorylation sites in the activation loop, thereby preventing their phosphorylation and subsequent signal transduction.

AvrAC contains an N-terminal domain (NTD) of unknown function, a leucine-rich-repeat domain (LRR), and a C-terminal region (CTR) harbouring the Fic domain (Fig. 4a). The homology of the AvrAC Fic domain with other Fic domains mainly resides in the C-terminal subdomain (Supplementary Fig. 11), in which the highly conserved HPFx(D/E)GN(G/K)R motif is responsible for nucleotide binding and catalysis^{22,23}. In particular, the invariant His residue

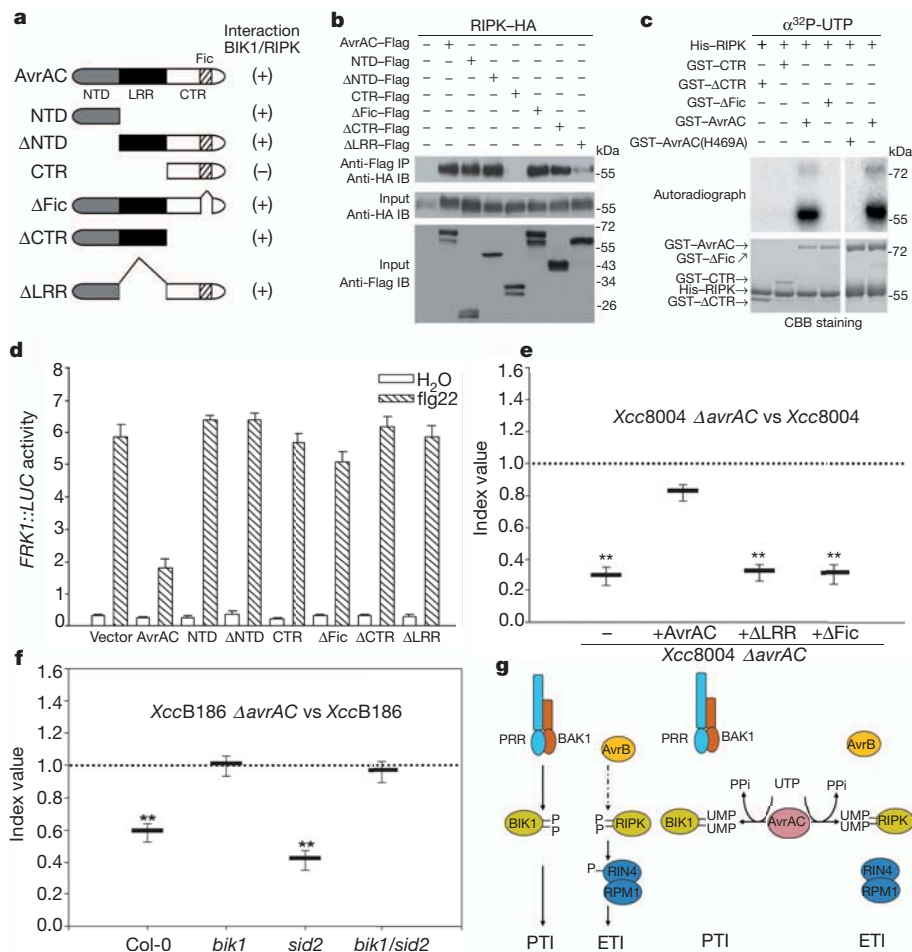


Figure 4 | AvrAC uridylyl transferase activity and BIK1 are required for AvrAC virulence in plants. **a**, Schematic diagram of AvrAC mutant constructs and their interaction with BIK1 and RIPK. The ability of these constructs to interact with BIK1 and RIPK is summarized on the right. **b**, The AvrAC NTD and LRR, but not CTR, interacted with RIPK in co-immunoprecipitation assays. **c**, In vitro uridylylation of RIPK by the indicated GST-AvrAC recombinant proteins. **d**, Inhibition of flg22-induced *FRK1::LUC* expression by AvrAC mutants (mean \pm s.d.; $n \geq 3$). **e**, The LRR domain and Fic motif are required for AvrAC virulence in cabbage plants. Competitive index assays were carried out on cabbage cultivar Zhonggan15 with *Xcc8004ΔavrAC* complemented with the indicated constructs (mean \pm s.d.; $n \geq 6$; $**P \leq 0.01$, Student's *t*-test). **f**, AvrAC

corresponding to AvrAC His 469 has a key role in the catalysis. AvrAC constructs lacking the Fic catalytic motif (Δ Fic) or CTR (Δ CTR) were fully capable of interacting with BIK1 and RIPK, whereas the AvrAC CTR alone was completely unable to interact with BIK1 and RIPK (Fig. 4b and Supplementary Fig. 12). Furthermore, the AvrAC NTD alone, the construct lacking NTD, and the construct lacking LRR were all able to interact with BIK1, indicating that NTD and LRR constitute the kinase-binding region. Removing the kinase-binding region or CTR, or substitution of His 469 to Ala, rendered AvrAC completely inactive in RIPK uridylylation *in vitro* (Fig. 4c), and induced a BIK1 mobility shift in protoplasts (Supplementary Figs 12 and 13). These results are consistent with the notion that the AvrAC Fic domain is responsible for catalysis, whereas NTD and LRR allow docking of substrate proteins.

We next determined the role of individual AvrAC domains in the inhibition of PTI signalling and virulence function. All the AvrAC deletion constructs were unable to inhibit *FRK1::LUC* expression, whereas the AvrAC(H469A) mutant was partially affected (Fig. 4d, Supplementary Figs 14 and 15). Only the wild-type, but not the truncated versions of *avrAC*, was able to promote *Xcc8004ΔavrAC* bacterial growth on Zhonggan15 plants (Fig. 4e). Together these

results indicated that the ability of AvrAC to inhibit immune signalling and enhance virulence not only required its binding to, but also UMP modification of, the substrate proteins. *Xcc8004* and *Xcc8004ΔavrAC* are weakly virulent on *Arabidopsis* plants⁹, preventing a test of AvrAC virulence function in *Arabidopsis*. We found that *Xcc* strain B186, which is highly virulent on *Arabidopsis* plants²⁴, contains AvrAC with a single amino acid substitution (AvrAC(T358A)). We then generated a *XccB186ΔavrAC* mutant and conducted competitive index assays in *Arabidopsis* mesophyll tissues. *XccB186ΔavrAC* grew to ~60% of *XccB186* in wild-type *Arabidopsis* plants, and the difference was statistically significant (Fig. 4f). In contrast, the two strains grew to nearly identical levels in *bik1* mutant plants. The *bik1* mutant is known to accumulate high levels of the defence hormone salicylate²⁵, which may potentially affect the virulence assay. We therefore carried out the assays on the *sid2* mutant, which has greatly reduced salicylate production. *XccB186ΔavrAC* grew to ~40% of *XccB186* on *sid2* plants (Fig. 4f). In contrast, the two strains grew to nearly identical levels in the *bik1/sid2* double-mutant plants, again indicating that *avrAC* promotes *XccB186* bacterial growth in a BIK1-dependent manner. Together these results demonstrated that BIK1 is a key target for AvrAC virulence activity.

results indicated that the ability of AvrAC to inhibit immune signalling and enhance virulence not only required its binding to, but also UMP modification of, the substrate proteins.

Xcc8004 and *Xcc8004ΔavrAC* are weakly virulent on *Arabidopsis* plants⁹, preventing a test of AvrAC virulence function in *Arabidopsis*. We found that *Xcc* strain B186, which is highly virulent on *Arabidopsis* plants²⁴, contains AvrAC with a single amino acid substitution (AvrAC(T358A)). We then generated a *XccB186ΔavrAC* mutant and conducted competitive index assays in *Arabidopsis* mesophyll tissues. *XccB186ΔavrAC* grew to ~60% of *XccB186* in wild-type *Arabidopsis* plants, and the difference was statistically significant (Fig. 4f). In contrast, the two strains grew to nearly identical levels in *bik1* mutant plants. The *bik1* mutant is known to accumulate high levels of the defence hormone salicylate²⁵, which may potentially affect the virulence assay. We therefore carried out the assays on the *sid2* mutant, which has greatly reduced salicylate production. *XccB186ΔavrAC* grew to ~40% of *XccB186* on *sid2* plants (Fig. 4f). In contrast, the two strains grew to nearly identical levels in the *bik1/sid2* double-mutant plants, again indicating that *avrAC* promotes *XccB186* bacterial growth in a BIK1-dependent manner. Together these results demonstrated that BIK1 is a key target for AvrAC virulence activity.

In this study we demonstrated that AvrAC inhibits both PTI and RPM1-specified ETI by targeting two related RLCKs, BIK1 and RIPK, as illustrated in the model (Fig. 4g). BIK1 and PBL1 are not required for PAMP-induced MAPK activation (Supplementary Fig. 16), suggesting that additional host targets account for the MAPK inhibition by AvrAC. Although *Xcc* is not known to trigger the RPM1-specific ETI in *Arabidopsis*, the *Xcc* bacterium carries XopAH (also known as AvrXccC), which is highly homologous to AvrB²⁶. It is formally possible that XopAH triggers ETI specified by an unknown resistance protein in some plants through RIPK. Inactivation of RIPK by AvrAC may allow the *Xcc* bacterium to inhibit such a potential ETI resistance in plants. AvrAC is the only Fic-domain effector among all sequenced phytopathogenic bacteria. However, the Fic and Doc (death on curing) domains contain a core structure, designated Fido, which is shared by AvrB and XopAH²⁷. AvrB and XopAH do not carry the catalytic His residue conserved in Fic and Doc proteins, and it remains to be determined whether they also possess phosphotransferase activities.

AvrAC uridylylates and masks the conserved phosphorylation sites in the activation loop of BIK1 and RIPK, preventing the activation of these kinases and impairing plant immunity. Interestingly, a similar strategy has been used by the *Yersinia* effector YopJ, which selectively acetylates serine and threonine residues in the activation loop of human MEK6, thereby blocking the phosphorylation of the target protein and signalling²⁸. Multiple *P. syringae* type III effectors are known to target plant protein kinases and inhibit PTI signalling by using diverse biochemical mechanisms⁵. The work described here illustrates a unique biochemical mechanism by which the *Xcc* bacterium combats the plant innate immune system.

METHODS SUMMARY

In vitro uridylylation assay. Affinity-purified recombinant His-RIPK or His-BIK1 protein was incubated with GST-AvrAC in an uridylylation reaction buffer supplied with α -³²P UTP. The products were separated on a SDS-PAGE gel and visualized by autoradiography.

Mapping UMP modification sites. Affinity-purified GST-AvrAC, His-BIK1, His-RIPK, BIK1-Flag and RIPK-Flag proteins were separated by SDS-PAGE. The corresponding protein bands were sliced and in-gel digested with trypsin overnight at 37 °C. The resulting peptides were extracted and analysed by nano-flow LC-MS/MS on a QSTAR XL mass spectrometer. Database searches were performed on a Mascot server with the following variable modifications: oxidation on methionine, carbamidomethylation on cysteine, UMP modification on serine, threonine or tyrosine, and AMP modification on serine, threonine or tyrosine.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 September 2011; accepted 17 February 2012.

Published online 15 April 2012.

1. Ausubel, F. M. Are innate immune signaling pathways in plants and animals conserved? *Nature Immunol.* **6**, 973–979 (2005).
2. Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803–814 (2006).
3. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
4. Boller, T. & Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **60**, 379–406 (2009).
5. Block, A. & Alfano, J. R. Plant targets for *Pseudomonas syringae* type III effectors: virulence targets or guarded decoys? *Curr. Opin. Microbiol.* **14**, 39–46 (2011).
6. Lu, D. *et al.* A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc. Natl Acad. Sci. USA* **107**, 496–501 (2010).

7. Zhang, J. *et al.* Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host Microbe* **7**, 290–301 (2010).
8. Liu, J., Elmore, J. M., Lin, Z.-J. D. & Coaker, G. A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe* **9**, 137–146 (2011).
9. Xu, R.-Q. *et al.* AvrAC_{Xcc8004}, a type III effector with a leucine-rich-repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *J. Bacteriol.* **190**, 343–355 (2008).
10. He, P. *et al.* Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. *Cell* **125**, 563–575 (2006).
11. Segonzac, C. & Zipfel, C. Activation of plant pattern-recognition receptors by bacteria. *Curr. Opin. Microbiol.* **14**, 54–61 (2011).
12. Chinchilla, D. *et al.* A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500 (2007).
13. Heese, A. *et al.* The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc. Natl Acad. Sci. USA* **104**, 12217–12222 (2007).
14. Chung, E.-H. *et al.* Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants. *Cell Host Microbe* **9**, 125–136 (2011).
15. Woolery, A. R., Luong, P., Broberg, C. A. & Orth, K. AMPylation: something old is new again. *Front. Microbiol.* **1**, 113 (2010).
16. Mattoo, S. *et al.* Comparative analysis of *Histophilus somni* IbpA with other FIC enzymes reveals differences in substrate and nucleotide specificities. *J. Biol. Chem.* **286**, 32834–32842 (2011).
17. Yarbrough, M. L. *et al.* AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* **323**, 269–272 (2009).
18. Worby, C. A. *et al.* The Fic domain: regulation of cell signaling by adenylylation. *Mol. Cell* **34**, 93–103 (2009).
19. Muller, M. P. *et al.* The *Legionella* effector protein DrrA AMPylates the membrane traffic regulator Rab1b. *Science* **329**, 946–949 (2010).
20. Mukherjee, S. *et al.* Modulation of Rab GTPase function by a protein phosphocholine transferase. *Nature* **477**, 103–106 (2011).
21. Laluk, K. *et al.* Biochemical and genetic requirements for function of the immune response regulator BOTRYTIS-INDUCED KINASE1 in plant growth, ethylene signaling, and PAMP-triggered immunity in *Arabidopsis*. *Plant Cell* **23**, 2831–2849 (2011).
22. Luong, P. *et al.* Kinetic and structural insights into the mechanism of AMPylation by VopS Fic domain. *J. Biol. Chem.* **285**, 20155–20163 (2010).
23. Xiao, J. *et al.* Structural basis of Fic mediated adenylylation. *Nature Struct. Mol. Biol.* **17**, 1004–1010 (2010).
24. Sun, W., Dunning, M. F., Pfund, C., Weingarten, R. & Bent, A. F. Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. *Plant Cell* **18**, 764–779 (2006).
25. Veronese, P. *et al.* The membrane-anchored BOTRYTIS-INDUCED KINASE1 plays distinct roles in *Arabidopsis* resistance to necrotrophic and biotrophic pathogens. *Plant Cell* **18**, 257–273 (2006).
26. White, F. F., Potnis, N., Jones, J. B. & Koebernik, R. The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* **10**, 749–766 (2009).
27. Kinch, L. N., Yarbrough, M. L., Orth, K. & Grishin, N. V. Fido, a novel AMPylation domain common to Fic, Doc, and AvrB. *PLoS ONE* **4**, e5818 (2009).
28. Mukherjee, S. *et al.* *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* **312**, 1211–1214 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank J. Chai for sharing plasmids before publication, A. Bent for the XccB186 strain, S. Y. He and F. White for helpful comments. J.-M.Z. was supported by grants from the Chinese Ministry of Science and Technology (2011CB100700; 2010CB835301). S.C. was supported by a grant from the Chinese Ministry of Science and Technology (2010CB835204). C.H. was supported by Funds from Hainan University.

Author Contributions S.C., C.H. and J.-M.Z. conceived and designed the experiments. F.F., F.Y., W.R., X.W. and J.Z. performed the experiments. F.F., F.Y., S.C., C.H. and J.-M.Z. analysed the data. F.F., S.C., C.H. and J.-M.Z. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.-M.Z. (zhoujianmin@nibs.ac.cn) or C.H. (czhe@hainu.edu.cn).

METHODS

Plants and bacterial strains. *Arabidopsis* plants were grown in a growth room at 23 °C and 70% relative humidity with a 10/14 h day/night light cycle for 5 weeks before protoplast isolation or bacterial inoculation. Cabbage cultivars Zhonggan15, Chungan45, Jingfeng1, Zhonggan12, Zhonggan8 were grown in greenhouses at 25 °C. The bacterial strains used in this study included *Xcc8004* (ref. 9), *Xcc8004ΔhrcV* (ref. 29), *Xcc* strain B186 (*XccB186*) (ref. 23), *Pst* strain DC3000, *PstΔhrcC* (ref. 30), *Pst(avrB)*, *Pst(avrRpt2)*, *Pst(avrPphB)* and *Agrobacterium* GV3101.

Constructs, generation of knockout strains and transgenic plants. To generate recombinant protein constructs, full-length *AvrAC* and truncated mutants were PCR-amplified from *Xcc8004* genomic DNA and inserted between *EcoRI* and *XhoI* sites of pGEX-6p-1 (Amersham Biosciences) to generate GST-*AvrAC*, GST-CTR (C-terminal region), GST-ΔCTR and GST-ΔFic. A GST-*AvrAC*(H469A) mutant construct was generated by site-directed mutagenesis. The coding sequences of *Arabidopsis* *RIPK*, *BIK1* and *PBL1* were amplified from cDNA and inserted into pET28a or pET28b (Novagen) to generate His-*RIPK*, His-*BIK1* and His-*PBL1*.

To generate constructs for protoplast transfection assay, *RIPK*, *AvrAC* and its derivatives were PCR-amplified and inserted between *XhoI* and *Csp45I* sites of pUC19-35S-Flag-RBS vector³¹ to generate *RIPK*-Flag, *AvrAC*-Flag, NTD-Flag, ΔNTD-Flag, CTR-Flag, ΔCTR-Flag, ΔFic-Flag, ΔLRR-Flag and *AvrAC*(H469A)-Flag. *RIPK* and *AvrAC* were cloned into *KpnI* and *Sall* sites of pUC19-35S-HA-RBS³¹ to generate *RIPK*-HA and *AvrAC*-HA, respectively. *BIK1*-Flag, *BIK1*-HA and *BIK1* mutants were described previously⁷.

To generate *avrAC* knockout mutants in *Xcc8004* and *XccB186*, an internal sequence corresponding to nucleotides 30–500 of the *avrAC* coding region was PCR-amplified from genomic DNA of *Xcc8004* and cloned into the suicide plasmid pK18mob³². The recombinant plasmid was transferred into the wild-type *Xcc* strains by triparental conjugation, and the single crossover mutants were confirmed by PCR and designated *Xcc8004ΔavrAC* and *XccB186ΔavrAC*. To complement the *Xcc8004ΔavrAC* mutant, the 400 bp upstream sequence of the *avrAC* coding region was PCR-amplified from the genomic DNA of *Xcc8004* and ligated between the *XbaI* and *HindIII* sites of pLAFR6 (ref. 9), generating an intermediate plasmid carrying *avrAC* promoter sequence. A full-length *avrAC*, *avrACΔLRR* or *avrACΔFic* fragments were PCR-amplified using the pUC19-35S-Flag-RBS plasmids containing the corresponding *avrAC* fragments as templates (see above) before integrated into the intermediate plasmid. The resulting constructs were introduced into the *Xcc8004ΔavrAC* strain by triparental conjugation.

The *avrAC* coding region was PCR-amplified from *Xcc8004* genomic DNA, ligated into a modified pER8 vector^{31,33}. The resulting clone containing *AvrAC*-Flag under the control of the oestrogen-inducible promoter was transformed into *Arabidopsis* (Col-0) by *Agrobacterium*-mediated transformation according to standard protocols. Transgenic plants were selected on Murashige and Skoog (MS) plates containing hygromycin. Two independent transgenic lines were selected for experiments. The transgenic plants were sprayed with 50 mM oestradiol in a 0.01% silwet L-77 solution for 12 h to induce *AvrAC* protein expression.

Dual reporter assay. *AvrAC* and its truncation mutants in the pUC19-35S-Flag-RBS vector were co-transfected with *FRK1::LUC* (firefly luciferase) and 35S::*RLUC* (Renilla luciferase) into *Arabidopsis* protoplasts. The protoplasts were incubated overnight under low light, treated with 1 μM flg22, 1 μM elf18, or 200 μg ml⁻¹ chitin (Sigma) for 3 h. Protein was then isolated, and LUC activity was recorded by using the Dual-Luciferase Reporter system (Promega) according to the manufacturer's instructions.

Oxidative burst. Oestradiol-induced leaves were sliced into 1 mm strips, incubated in 200 μl water in a 96-well plate for 12 h before the addition of 1 μM flg22 in 200 μl reaction buffer supplement with 20 mM luminol and 1 μg horseradish peroxidase (Sigma). Luminescence was recorded with a Luminometer (Promega) for 20 min.

MAPK activity assay. Two-week-old wild-type and transgenic seedlings grown on 1/2 MS medium were sprayed with 10 μM flg22 or water containing 0.02% Silwet L-77 for 10 min and frozen in liquid nitrogen. The seedlings were homogenized in an extraction buffer containing 50 mM HEPES-KOH (pH 7.5), 150 mM KCl, 1 mM EDTA, 0.2% Triton-X 100, 1 mM DTT, complete protease inhibitors (Roche), and phosphatase inhibitors (Fisher). The protein concentration was determined using a Bio-Rad Bradford protein assay kit, and equal amounts of total protein were electrophoresed on 10% SDS-PAGE. An anti-pERK antibody (no. 4370S, Cell Signaling) was used to determine phosphorylation state of MPK3, MPK4 and MPK6 in an immunoblot. A duplicate blot was reacted with anti-MPK3 antibodies (Sigma) to determine the amount of total MPK3. For MAPK activation by MKK5^{DD}, protoplasts prepared from wild-type plants were transfected with the indicated plasmids, and protein was isolated 12 h later for

immunoblot analyses. The identities of phosphorylated MAPKs were labelled as described previously^{34,35}.

Bacterial growth and hypersensitive response assays. Five-week-old *Arabidopsis* leaves were infiltrated with the indicated *Xcc* or *Pst* bacteria at 10⁶ c.f.u. ml⁻¹, and the bacterial population in the leaf was counted at the indicated times. Each data point consisted of at least four replicates. For the hypersensitive response assay, wild-type plants were first infiltrated with *Xcc8004* or *Xcc8004ΔavrAC* mutant bacteria at 10⁷ c.f.u. ml⁻¹ 1 day before infiltrating 5 × 10⁷ c.f.u. ml⁻¹ *Pst* or *Pst(avrB)*. Plants were photographed 6 h after the second inoculation.

A competitive index assay³⁶ was used to compare virulence between wild-type and *ΔavrAC* mutant strains of *Xcc*. Briefly, equal amounts (c.f.u.) of wild-type and *ΔavrAC* mutant *Xcc* bacteria were mixed and inoculated into plants, and leaf bacterial number for each strain was determined by plating on NYG agar medium supplemented with appropriate antibiotics. The ratio between the mutant and wild-type bacteria (competitive index) was calculated at the desired times. For virulence assay on cabbage, fully expanded leaves of 4-week-old plants were clipping inoculated³⁷ with a mixture of *Xcc8004* and *Xcc8004ΔavrAC* bacteria at an equal concentration of 10⁸ c.f.u. ml⁻¹, maintained at high humidity for 48 h under a transparent dome. Leaf bacterial number and competitive index were determined 6 days after inoculation. For virulence assay on *Arabidopsis* plants, 5-week-old *Arabidopsis* leaves were infiltrated with a mixture of wild-type *XccB186* and *XccB186ΔavrAC* mutant bacteria at an equal concentration of 10⁶ c.f.u. ml⁻¹. Leaf bacterial number and competitive index were determined 3 days after inoculation.

CyaA secretion assay. The plasmid pCPP3221 containing an *AvrPto*-Cya fusion sequence was introduced into the wild-type *Pst* (DC3000) and its *ΔhrcC* mutant derivative. The derived strains were inoculated at 1 × 10⁸ c.f.u. ml⁻¹ into *Arabidopsis* leaves that had been pre-treated with 10⁷ c.f.u. ml⁻¹ *Xcc8004* or *Xcc8004ΔavrAC* for 24 h. Amounts of *AvrPto*-Cya secretion were determined as described previously³⁸. Briefly, leaf samples were ground in liquid nitrogen and resuspended in 0.1 M HCl. The total protein content of each sample was determined using the Bradford assay (Bio-Rad). A direct cAMP (cAMP) immunoassay kit (Sigma) was used to measure cAMP levels following the manufacturer's instruction.

Recombinant proteins expression and GST pull-down assay. The recombinant proteins were affinity purified following the manufacturer's instruction. For pull-down assays, 10 μg each of His-*RIPK*, His-*BIK1* and GST-*AvrAC* were incubated on a rotator with glutathione agarose beads (GE Healthcare) in a buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM DTT for 3 h, washed five times with a buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM DTT. The bound protein was eluted with 15 mM GSH and immunoblotted with a mouse anti-His monoclonal antibody (Sigma-Aldrich).

Co-immunoprecipitation assay. The protoplasts were transfected with the indicated constructs, incubated for 12 h, treated with either H₂O or 1 μM flg22 for 3–10 min, and total protein was isolated with an extraction buffer (see MAPK activity assay). For anti-Flag immunoprecipitation, total protein was incubated with an agarose-conjugated anti-Flag antibody (Sigma-Aldrich) for 4 h, washed six times with a buffer containing 50 mM HEPES-KOH (pH 7.5), 150 mM KCl, 1 mM EDTA, 0.2% Triton-X 100, 1 mM DTT, and the bound protein was eluted with 0.5 mg ml⁻¹ 3×Flag peptide. For anti-HA immunoprecipitation, total protein was pre-cleared with protein A agarose (Millipore) for 1 h, followed by precipitation with 2 μg anti-HA antibody (TianGen) together with protein A agarose for 4 h. Immunoprecipitates were separated by a 10% NuPAGE gel (Invitrogen) and detected by anti-HA or anti-Flag immunoblot.

In vitro kinase assay. His-*RIPK*, His-*BIK1* and GST-*AvrAC* were co-expressed in *E. coli* (strain BL21), and the His-tagged protein was purified by Ni-NTA affinity chromatography. Kinase assays were performed in 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 100 μM ATP, 1 mM DTT, γ-³²P ATP (5 μCi), and 3 μg recombinant protein in a total volume of 20 μl. Reactions were allowed to proceed for 30 min at 30 °C and terminated by adding 5× protein loading buffer and boiling for 5 min at 100 °C. The proteins were separated on a 12% SDS-PAGE gel, and incorporated radiolabel was visualized by autoradiography. As an equal loading control for proteins used in the kinase reactions, the SDS-PAGE gels were stained by Coomassie blue.

BIK1 phosphorylation and migration shift assay. *Arabidopsis* protoplasts were transfected with HA-tagged *BIK1* or *BIK1* mutants⁷ together with *AvrAC*-Flag or truncation mutants of *AvrAC*, treated with 1 μM flg22 and total protein was extracted at various time points. Samples were separated by 10% SDS-PAGE gels followed by anti-HA immunoblot. For phosphatase treatment, total protein from protoplasts or affinity-purified recombinant protein was treated with λ protein phosphatase (New England Biolabs) according to the manufacturer's instructions.

In vitro uridylylation assay. GST-tagged AvrAC derivative proteins, His-tagged RIPK, BIK1 and PBL1 proteins were affinity purified and directly used for the uridylylation assay. The GST-PBS1 recombinant protein (a gift from J. Chai) was cleaved with PreScission Protease (GE) to remove GST before the enzymatic assay. Approximately 400 ng of GST-AvrAC, GST-ΔCTR, GST-CTR, GST-ΔFic or GST-AvrAC(H469A) were incubated with 2 μg of His-RIPK, His-BIK1, PBS1, His-PBL1, or His-tagged kinase domain of FLS2 or BAK1 (ref. 39) in 20 μl reaction buffer containing 25 mM Tris-HCl (pH 7.5), 25 mM MgCl₂, 500 μM UTP, 1 mM DTT, α-³²P UTP (5 μCi) for 30 min at 30 °C. Reactions were stopped by addition of loading buffer. The products were separated on 12% NuPAGE gels (Invitrogen) and visualized by autoradiography.

Mass measurement of the recombinant AvrAC protein. An affinity-purified GST-AvrAC protein solution was loaded into an analytical capillary column (75 μm × 2 cm) packed with Poros 20 R2 packing material (Applied Biosystems). An Agilent 1100 binary pump was used to generate HPLC gradient as follows: 0–100% B in 60 min (A = 0.1 M acetic acid in water; B = 0.1 M acetic acid/70% acetonitrile). The eluted proteins were sprayed into a QSTAR XL mass spectrometer (AB Sciex). The spray voltage was set at 2,100 V and the data was acquired in MS mode. The protein charge envelope from the raw spectrum was de-convoluted into non-charged form by the BioAnalyst software provided by the manufacturer.

Mapping UMP modification sites. Affinity-purified GST-AvrAC, His-BIK1, His-RIPK, BIK1-Flag and RIPK-Flag proteins were separated by SDS-PAGE. Protein bands on the SDS-PAGE gel were de-stained, and then reduced in 10 mM DTT at 56 °C for 30 min followed by alkylation in 55 mM iodoacetamide at dark for 1 h. After that the protein bands were in-gel digested with sequencing grade trypsin (10 ng μl⁻¹ trypsin, 50 mM ammonium bicarbonate, pH 8.0) overnight at 37 °C. Peptides were extracted with 5% formic acid/50% acetonitrile and 0.1% formic acid/75% acetonitrile sequentially and then concentrated to ~20 μl. The extracted peptides were separated by an analytical capillary column (50 μm × 10 cm) packed with 5 μm spherical C18 reversed phase material (YMC). An Agilent 1100 binary pump was used to generate the HPLC gradient as follows: 0–5% B in 5 min, 5–40% B in 25 min, 40–100% B in 15 min (A = 0.1 M acetic acid in water; B = 0.1 M acetic acid/70% acetonitrile). The eluted peptides were sprayed into a QSTAR XL mass spectrometer equipped with a nano-ESI ion source. The mass

spectrometer was operated in information-dependent mode with one MS scan followed by three MS/MS scans for each cycle. Database searches were performed on an in-house Mascot server (Matrix Science) and the following variable modifications were included: oxidation on methionine, carbamidomethylation on cysteine, UMP modification on serine or threonine residues, AMP modification on serine, threonine or tyrosine residues.

29. Wang, L. F., Tang, X. Y. & He, C. Z. The bifunctional effector AvrXccC of *Xanthomonas campestris* pv. *campestris* requires plasma membrane-anchoring for host recognition. *Mol. Plant Pathol.* **8**, 491–501 (2007).
30. Yuan, J. & He, S. Y. The *Pseudomonas syringae* Hrp regulation and secretion system controls the production and secretion of multiple extracellular proteins. *J. Bacteriol.* **178**, 6399–6402 (1996).
31. Li, X. et al. Flagellin induces innate immunity in nonhost interactions that is suppressed by *Pseudomonas syringae* effectors. *Proc. Natl Acad. Sci. USA* **102**, 12990–12995 (2005).
32. Schäfer, A. et al. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73 (1994).
33. Zuo, J., Niu, Q. W. & Chua, N. H. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**, 265–273 (2000).
34. Zhang, J. et al. A *Pseudomonas syringae* effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell Host Microbe* **1**, 175–185 (2007).
35. Schwessinger, B. et al. Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genet.* **7**, e1002046 (2011).
36. Macho, A. P. et al. Competitive index in mixed infections: a sensitive and accurate assay for the genetic analysis of *Pseudomonas syringae*–plant interactions. *Mol. Plant Pathol.* **8**, 437–450 (2007).
37. Dow, J. M. et al. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl Acad. Sci. USA* **100**, 10995–11000 (2003).
38. Schechter, L. M. et al. *Pseudomonas syringae* type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *J. Bacteriol.* **186**, 543–555 (2004).
39. Cheng, W. et al. The AvrPtoB-BAK1 complex reveals two structurally similar kinase-interacting domains in a single type III effector. *Cell Host Microbe* **10**, 616–626 (2011).

Spin–orbital separation in the quasi–one–dimensional Mott insulator Sr_2CuO_3

J. Schlappa^{1,2}, K. Wohlfeld³, K. J. Zhou^{1†}, M. Mourigal⁴, M. W. Haverkort⁵, V. N. Strocov¹, L. Hozoi³, C. Monney¹, S. Nishimoto³, S. Singh^{6†}, A. Revcolevschi⁶, J.-S. Caux⁷, L. Patthey^{1,8}, H. M. Rønnow⁴, J. van den Brink³ & T. Schmitt¹

When viewed as an elementary particle, the electron has spin and charge. When binding to the atomic nucleus, it also acquires an angular momentum quantum number corresponding to the quantized atomic orbital it occupies. Even if electrons in solids form bands and delocalize from the nuclei, in Mott insulators they retain their three fundamental quantum numbers: spin, charge and orbital¹. The hallmark of one-dimensional physics is a breaking up of the elementary electron into its separate degrees of freedom². The separation of the electron into independent quasi-particles that carry either spin (spinons) or charge (holons) was first observed fifteen years ago³. Here we report observation of the separation of the orbital degree of freedom (orbiton) using resonant inelastic X-ray scattering on the one-dimensional Mott insulator Sr_2CuO_3 . We resolve an orbiton separating itself from spinons and propagating through the lattice as a distinct quasi-particle with a substantial dispersion in energy over momentum, of about 0.2 electronvolts, over nearly one Brillouin zone.

It was pointed out in the 1970s that in a solid not only the charge and spin of electrons can become ordered—leading to magnetism—but also the electrons' orbital degree of freedom¹. This observation sparked a field that has gone on to produce a number of important results. Although a physical electron combines spin, charge and orbital, theoretically an electron can be considered a bound state of the three independent, fundamental quasi-particles: a spinon, carrying the electron's spin; a holon (or chargon), carrying its charge; and an orbiton, carrying its orbital degree of freedom.

A remarkable and fundamental property of one-dimensional (1D) systems is that electronic excitations break up into deconfined spinons and holons. This was predicted decades ago (ref. 2 and references therein) and confirmed in the mid 1990s by angle-resolved photoemission spectroscopy experiments^{3–5}. The spin–charge separation is an example of particle fractionalization, a phenomenon in which the quantum numbers of quasi-particles are not multiples of those of the elementary particle, but fractions. This effect is one of the most unusual manifestations of collective quantum physics of interacting particles and is a profound concept that has found its way into a number of theories, for example that describing high-temperature superconductivity in copper oxides^{6,7}.

To search for the further fractionalization of the electron, we consider the excitation of a copper orbital degree of freedom in the antiferromagnetic spin-chain compound Sr_2CuO_3 . The spin–orbital separation process that we are looking for is analogous to the spin–charge separation mechanism (Fig. 1b). The latter occurs, for instance, when an electron is annihilated, removing a single spin and leaving behind a hole in the antiferromagnetic chain. This hole can start to propagate freely only after exciting one spinon (a domain wall in the antiferromagnetic chain). Subsequently, the spinon can delocalize and

separate itself completely from the holon. When instead of creating a hole, as typically is done in a photoemission experiment, an electron is excited from one copper 3d orbital to another, the phenomenon of spin–orbital separation can in principle occur (Fig. 1a). The orbiton created in this manner may also deconfine after exciting a spinon, thus splitting the electron into its orbital and spin degrees of freedom⁸.

Here we use high-resolution resonant inelastic X-ray scattering (RIXS) to search experimentally for spin–orbital separation in the quasi-1D copper oxide Sr_2CuO_3 (for material details, see Supplementary Information, section 1). We observe deconfinement of the spinon and orbiton during orbital excitation from the ground-state copper 3d $x^2 - y^2$ orbital to an excited copper 3d xy or xz orbital (Fig. 1c–e). For simplicity, we will from henceforth use the so-called 'hole' language: although nominally there are nine electrons in the 3d orbitals of the Cu^{2+} ion in Sr_2CuO_3 , by using the 'electron–hole' transformation we can map the problem onto an effective system with one particle occupying a single 3d orbital (Supplementary Information, section 2).

We measure an orbiton dispersion that is almost as large as the dispersion of the two-spinon continuum at low energies. As for spin–charge separation^{3–5}, the orbiton dispersion has periodicity π (see Fig. 1c and see discussion below), which indicates the presence of an orbiton liberated from the spinon.

We measured the orbital excitations of Sr_2CuO_3 using RIXS at the L_3 edge of the copper ion. RIXS is a second-order scattering technique and can excite transitions between the copper 3d states of different symmetry (orbital excitations), owing to the involvement of two subsequent electric dipole transitions^{9,10} (Supplementary Information, section 3). With the unique capability of RIXS also to probe spin excitations^{11–13} and to vary the photon momentum transfer, the dispersion of orbital and spin excitations can be mapped out across the first Brillouin zone^{11–16}. The experiments were carried out at the ADRESS beamline of the Swiss Light Source at the Paul Scherrer Institut^{17,18}.

For fixed momentum transfer, \mathbf{q} , along the chains, peaks in the RIXS spectrum at constant energy transfer reveal the presence of charge-neutral elementary excitations and are visible in the RIXS intensity map of Sr_2CuO_3 across the copper L_3 edge in Fig. 2a. The spectrum for which the incident energy was precisely tuned to the resonance maximum of the absorption spectrum is shown in Fig. 2b. In both plots, the excitations of the spin, orbital and charge degrees of freedom are indicated. The momentum dependence and, in particular, the dispersion of the spin and orbital excitations (Fig. 1c and Supplementary Fig. 2a) are indicative of their collective nature.

For energy transfers of up to ~ 0.8 eV purely magnetic excitations are present, but the spectrum between ~ 1.5 and ~ 3.5 eV corresponds to excitations from the copper 3d $x^2 - y^2$ ground state to orbitals of xy , xz/yz and $3z^2 - r^2$ symmetry (Fig. 1d, e). These peaks correspond to

¹Paul Scherrer Institut, Swiss Light Source, CH-5232 Villigen PSI, Switzerland. ²Institut Methoden und Instrumentierung der Forschung mit Synchrotronstrahlung G-12, Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, D-12489 Berlin, Germany. ³Institute for Theoretical Solid State Physics, IFW Dresden, Helmholtzstrasse 20, 01069 Dresden, Germany. ⁴Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. ⁵Max Planck Institute for Solid State Research, D-70569 Stuttgart, Germany. ⁶ICMMO - UMR 8182 - Bâtiment 410, Université Paris-Sud 11, 91405 Orsay Cedex, France. ⁷Institute for Theoretical Physics, University of Amsterdam, Science Park 904, Postbus 94485, 1090 GL Amsterdam, The Netherlands. ⁸SwissFEL, CH-5232 Villigen PSI, Switzerland. [†]Present addresses: Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK (K.J.Z.); Indian Institute of Science Education and Research, 900 NCL Innovation Park, Pashan, 411008 Pune, India (S.S.).

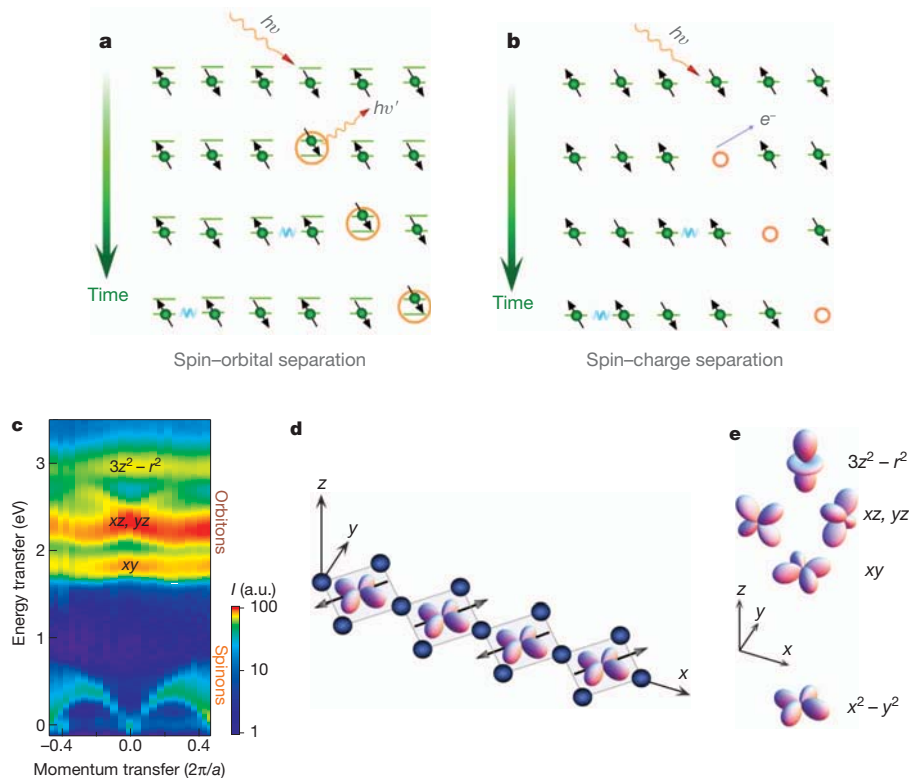
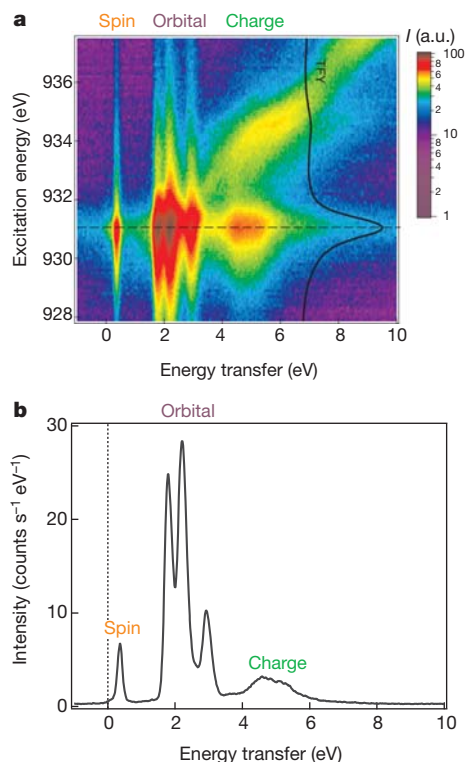


Figure 1 | Spin-orbital separation process in an antiferromagnetic spin chain, emerging after exciting an orbital. **a, b,** Sketches of spin-orbital separation (**a**) and spin-charge separation (**b**), generated in processes of RIXS and angle-resolved photoemission spectroscopy, respectively. The spin is represented by the arrow and the charge is represented by the green dot. In **a**, the lower and upper lines represent the ground-state and excited-state orbitals, respectively. **c**, RIXS intensity map of the dispersing spin and orbital excitations in Sr_2CuO_3 as functions of photon momentum transfer along the

chains and photon energy transfer (for details, see main text and Supplementary Information, section 3). **a**, $[100]$ lattice constant; a.u., arbitrary units. Data obtained at $\Psi = 130^\circ$. **d**, Geometry of the CuO_3 chain, with the ground-state copper $3d\ x^2 - y^2$ orbitals in the middle of each plaquette and oxygen sites at the plaquette corners. **e**, Orbital symmetries of $x^2 - y^2$ and excited $3d$ orbitals. In panels **c**–**e**, ‘hole’ language is used (Supplementary Information, section 2).



orbital excitations (called also d – d excitations), and not, for example, to charge transfer excitations, the intensity of which is non-zero for energies up to ~ 6 eV but here for the case of L -edge RIXS is at least an order of magnitude lower¹⁰. The orbital assignment of these excitations was unambiguously verified by comparing their energy at $\mathbf{q} = 0$ with *ab initio* quantum chemistry cluster calculations¹⁹ (Supplementary Information, section 5, for detailed results).

Zooming into the ‘magnetic’ part of Fig. 1c, between 0 and 0.8 eV in energy transfer, reveals strongly dispersing spin excitations. At the lower boundary the dispersion has period π , and at the continuum (upper boundary) it has period 2π (Fig. 3a). These RIXS data agree very well with recent inelastic neutron scattering studies on Sr_2CuO_3 (ref. 20). The simultaneous presence in the spectrum of a lower edge with period π and an upper one with period 2π indicates directly that in the spin chain the magnetic excitations with spin 1 break up and fractionalize into two-spinon (and higher-order) excitations that make up a continuum²⁰, with each spinon having spin $1/2$. These spectra confirm that RIXS for magnetic excitations probes the well-known spin dynamical structure factor as theoretically predicted^{11–13}, in agreement with recent studies on TiOCl (ref. 21). The excellent statistics of

Figure 2 | Energy dependence of elementary excitations in Sr_2CuO_3 observed with RIXS at the copper L_3 -edge resonance. **a**, RIXS map of Sr_2CuO_3 (log scale) as a function of photon excitation energy (left axis) and energy transfer (bottom axis). The superimposed black curve shows the total fluorescence yield X-ray absorption spectrum. The dashed line marks the maximum energy of the copper L_3 -edge resonance. **b**, RIXS line spectrum measured at the resonance maximum (along the dashed line in **a**). All data was obtained at a scattering angle of $\Psi = 90^\circ$ and for momentum transfer along the chain of $q = 0.189 \times 2\pi/a$ (Methods).

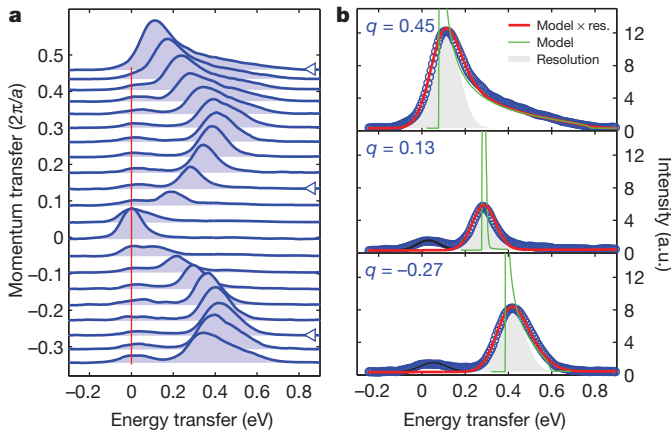


Figure 3 | Dispersion of magnetic excitations: experimental data and simulation. **a**, RIXS energy transfer spectra of the magnetic two-spinon continuum (see the spinon region in Fig. 1c). Data obtained at $\Psi = 130^\circ$. **b**, Fit (red line) of the experimental data (dots) with the two- and four-spinon dynamical structure factor²² (green line), convolved with a Gaussian distribution (light-grey shaded region) to account for the total instrumental resolution of 140 meV. The low-energy peak at around 50 meV (black line) is elastic and phonon scattering. The values of momentum transfer for the selected spectra (shown in units of $2\pi/a$) are indicated by arrows in **a**.

the data further allow for a direct comparison of the RIXS line shapes with the exact two- and four-spinon dynamical structure factor of the spin-1/2 Heisenberg chain. In Fig. 3b, we show the fits for three selected momentum transfer values using the exact two- and four-spinon dynamical structure factor, $S(\mathbf{q}, \omega)$ (where \mathbf{q} and ω stand for the momentum and energy of the created excitation, respectively) in the representation of ref. 22. The obtained exchange coupling, $J = 249$ meV, is in very good agreement with the value obtained from inelastic neutron scattering data²⁰.

Having unambiguously identified the fractionalized spinon excitations in the low-energy sector, we now concentrate on the orbital excitations spectrum in Fig. 4. We find that these are strongly momentum dependent and have a novel, distinct dispersion. This proves that the orbital excitations observed here are of collective nature. The xz excitation has the largest dispersion, of ~ 0.2 eV, and has a spectrum containing two peculiar components: a lower branch dispersing with periodicity π and, above that, an incoherent spectrum with a double-oval shape. This spectrum is strikingly similar to seemingly unrelated angle-resolved photoemission spectra of 1D copper oxides, which evidence spin–charge separation (see, for example, fig. 3 of ref. 3). This is an indication that the observed orbital dispersion is related to an analogous separation of degrees of freedom.

To test this conjecture, we derived a microscopic model that describes the spin–orbital interactions in Sr_2CuO_3 (Methods). The low-energy Kugel–Khomskii Hamiltonian for this was obtained^{8,23}

from the charge transfer model of Sr_2CuO_3 in ref. 24. The crucial part of the Hamiltonian, responsible for the xz orbital propagation, is

$$H = -J_O \sum_{j,\sigma} (c_{j\sigma}^+ c_{j+1,\sigma} + \text{h.c.}) + J \sum_j \mathbf{S}_j \cdot \mathbf{S}_{j+1} + E_O \sum_j (1 - n_j) \quad (1)$$

where J and J_O are respectively the spin and orbital exchange constants, both of which are fixed by the charge transfer model²⁴; E_O is the xz orbital on-site energy; $c_{j\sigma}^+$ and $c_{j\sigma}$ are operators that respectively annihilate and create particles in xz orbital with spin σ ; \mathbf{S}_j is the spin of the particle in $x^2 - y^2$ orbital; $1 - n_j$ counts the number of particles in xz orbital; and ‘h.c.’ denotes Hermitian conjugate. The first term in the Hamiltonian describes the propagation of orbital excitations through the lattice. The second term represents the usual Heisenberg interaction between spins, which vanishes on bonds where an orbital excitation is present (Methods).

From purely a mathematical standpoint, equation (1) is identical to the 1D t – J model—describing the hopping of holes doped into an antiferromagnetic spin chain in which the spins interact via superexchange J —with the orbital superexchange, J_O , taking the place of t , the hole-hopping amplitude in the t – J model. This implies that the propagation of a single orbital excitation in the J_O – J model above is equivalent to the propagation of a single hole in the 1D t – J Hamiltonian. Because in the 1D t – J model a hole breaks up into a free spinon and holon, in the J_O – J model the orbital excitation will separate into a free spinon and orbiton.

Before calculating RIXS spectral functions and quantifying the separation of spin and orbital degrees of freedom in Sr_2CuO_3 , it is instructive to consider the overall features of the momentum-dependent RIXS spectra within a slave-boson picture in terms of an orbital–spin separation ansatz. This ansatz is analogous to the mean-field slave-boson picture for charge–spin separation^{25,26}, which stipulates that the hole spectral function is given by the convolution of the spectral function of a free holon and a free spinon. In orbital–spin separation ansatz, free orbitons and spinons, created by the operators o_i^+ and $s_{i\sigma}^+$, form the particle creation operator $c_{i\sigma}^+ = s_{i\sigma}^+ o_i$. On applying this formalism to the Hamiltonian in equation (1), the energies of non-interacting orbitons and spinons with momentum k become $\varepsilon_O(k) = E_O - 2J_O \cos(k)$ and $\varepsilon_S(k) = -J \cos(k)$, respectively. In this ansatz, the spectral function for the orbital excitation produced in RIXS, $c_{q\sigma}^+ = \sum_k o_{q-k} s_{\sigma k}^+$ (where q , the x component of \mathbf{q} , is here the momentum transfer along the CuO_3 chain direction), is determined by a convolution of these two dispersions. The resulting excitation continuum is shown in Fig. 4b and has three defining features^{25,26}: a prominent lower edge with energy, $E(q) = E_O - 2J_O |\sin(q)|$, that is entirely determined by the dispersion of the orbiton; a distinctive upper orbiton edge at $E(q) = E_O + 2J_O |\sin(q)|$; and a broader upper limit of the orbiton–spinon continuum at $F(q) = E_O + \sqrt{J^2 + 4J_O^2 - 4JJ_O \cos(q)}$. These dispersion curves are directly compared to our RIXS experiments in

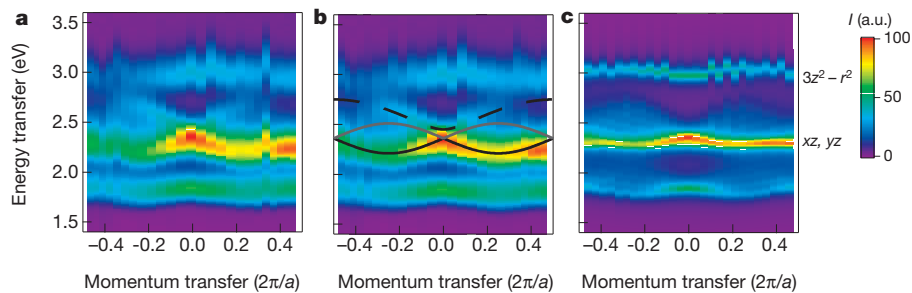


Figure 4 | Dispersion of orbital excitations: comparison between experiment and *ab initio* calculations. **a**, **c**, Comparison between the orbital region in Fig. 1c (**a**, left) and theory, using the J_O – J model (**c**, right). **b**, Lower and upper edges of the orbiton dispersion (solid black and grey lines) and the

upper edge of the spinon–orbiton continuum (dashed line), as calculated using the orbital–spin separation ansatz. See main text and Supplementary Information for further details.

Fig. 4b. The measured spectral weight that accumulates at the first two lower-energy features does so with periodicity π , which is a direct consequence of spin–orbital separation.

To quantify the spectral weights within the orbiton–spinon continuum, we have calculated for the J_O – J Hamiltonian the orbital excitation Green’s function by exactly diagonalizing the Hamiltonian on 28 lattice sites (as in the spin–charge separation studies³, finite-size effects are negligible (they are estimated at ~ 0.01 eV)). From this, we calculate the RIXS spectrum following ref. 13, that is, by expressing the RIXS amplitude as a product of the single-ion local RIXS effective operator and the orbital excitation Green’s function (Methods).

We find excellent agreement between theory and experiment (Fig. 4): both the sine-like xz orbiton dispersion and the xz spinon–orbiton continuum (the ‘double-oval’ incoherent spectrum), which are the hallmarks of spin–orbital separation, are present in the theory. The calculations also show that, in contrast to the xz orbital, the xy orbital has a small dispersion and that the excitations of both the $3z^2 - r^2$ orbital and the yz orbital are dispersionless, as is observed experimentally. This is an independent merit of the model because no fitting of dispersions to experimental data is involved.

The large orbiton dispersion observed in this study is the key feature that distinguishes Sr_2CuO_3 from other systems with orbital excitations^{27–29}, and relies on the 1D character of Sr_2CuO_3 . In a system of higher dimensionality, orbitons interact with magnetic excitations, which tend to slow them down and thus reduce their dispersion. In one dimension, orbitons can avoid these renormalization effects by means of spin–orbital separation.

METHODS SUMMARY

We applied the technique of high-resolution RIXS with the incident photon energy tuned to the L_3 edge ($2p_{3/2} \rightarrow 3d$ resonance) of the copper ion (around 931 eV) and a total experimental resolution of 140 meV. The experiments were performed at the ADRESS beamline of the Swiss Light Source at the Paul Scherrer Institut^{17,18}. The RIXS spectrometer was located at a fixed scattering angle of either $\Psi = 90^\circ$ or 130° and was collecting signal within the solid angle of $(19.3 \times 3.3) \text{ mrad}^2$ (horizontal \times vertical detection direction; for a sketch of the experimental geometry, see Supplementary Fig. 1).

Single-crystal samples of Sr_2CuO_3 were grown by the floating-zone method and freshly cleaved before the RIXS experiment. The surface normal to the sample, [010], and the propagation direction of the chains, [100], were oriented parallel to the scattering plane. The sample was cooled with a helium-flow cryostat to 14 K during the measurements. Incident photons were linearly polarized in the scattering plane (π -orientation). The momentum transfer along the chains, q , was varied by changing the incidence angle in steps of $5 \pm 1^\circ$.

Theoretical modelling was done by deriving a Kugel’–Khomskii Hamiltonian from the charge transfer model and then writing it in a representation similar to the one used for the t – J model. Using this model, we calculated the spectral functions for the case of a single orbital excitation in the antiferromagnetic background by solving the Hamiltonian numerically on 28 sites. Such calculations were repeated separately for the two distinct dispersive orbital excitations, whereas for the non-dispersive excitations we confirmed that the parameters of the model did not allow for a dispersion larger than the experimental resolution. Finally, the RIXS cross-section was calculated by multiplying the calculated spectral functions for the orbital excitations by the local RIXS form factors.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2011; accepted 17 February 2012.

Published online 18 April 2012.

1. Kugel’, K. I. & Khomskii, D. I. The Jahn–Teller effect and magnetism: transition metal compounds. *Sov. Phys. Usp.* **25**, 231–256 (1982).
2. Giamarchi, T. *Quantum Physics in One Dimension* (Clarendon Press, 2004).
3. Kim, C. *et al.* Observation of spin-charge separation in one-dimensional SrCuO_2 . *Phys. Rev. Lett.* **77**, 4054–4057 (1996).
4. Fujisawa, H. *et al.* Angle-resolved photoemission study of Sr_2CuO_3 . *Phys. Rev. B* **59**, 7358–7361 (1999).
5. Kim, B. J. *et al.* Distinct spinon and holon dispersions in photoemission spectral functions from one-dimensional SrCuO_2 . *Nature Phys.* **2**, 397–401 (2006).
6. Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).

7. Senthil, T. & Fisher, M. P. A. Fractionalization in the cuprates: detecting the topological order. *Phys. Rev. Lett.* **86**, 292–295 (2001).
8. Wohlfeld, K., Daghofer, M., Nishimoto, S., Khaliullin, G. & van den Brink, J. Intrinsic coupling of orbital excitations to spin fluctuations in Mott insulators. *Phys. Rev. Lett.* **107**, 147201 (2011).
9. Ament, L., van Veenendaal, M., Devereaux, T., Hill, J. P. & van den Brink, J. Resonant inelastic X-ray scattering studies of elementary excitations. *Rev. Mod. Phys.* **83**, 705–767 (2011).
10. Moretti Sala, M. *et al.* Energy and symmetry of dd excitations in undoped layered cuprates measured by $\text{Cu } L_3$ resonant inelastic x-ray scattering. *N. J. Phys.* **13**, 043026 (2011).
11. Ament, L. J. P., Ghiringhelli, G., Moretti Sala, M., Braicovich, L., & van den Brink, J. Theoretical demonstration of how the dispersion of magnetic excitations in cuprate compounds can be determined using resonant inelastic x-ray scattering. *Phys. Rev. Lett.* **103**, 117003 (2009).
12. Luo, J., Trammell, G. T. & Hannon, J. P. Scattering operator for elastic and inelastic resonant x-ray scattering. *Phys. Rev. Lett.* **71**, 287–290 (1993).
13. Haverkort, M. W. Theory of resonant inelastic X-ray scattering by collective magnetic excitations. *Phys. Rev. Lett.* **105**, 167404 (2010).
14. Schlappa, J. *et al.* Collective magnetic excitations in the spin ladder $\text{Sr}_{1.4}\text{Cu}_{2.4}\text{O}_{4.1}$ Measured using high-resolution resonant inelastic X-Ray scattering. *Phys. Rev. Lett.* **103**, 047401 (2009).
15. Braicovich, L. *et al.* Magnetic excitations and phase separation in the underdoped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ superconductor measured by resonant inelastic X-ray scattering. *Phys. Rev. Lett.* **104**, 077002 (2010).
16. Le Tacon, M. *et al.* Intense paramagnon excitations in a large family of high-temperature superconductors. *Nature Phys.* **7**, 725–730 (2011).
17. Strocov, V. N. *et al.* High-resolution soft X-ray beamline ADRESS at the Swiss Light Source for resonant inelastic X-ray scattering and angle-resolved photoelectron spectroscopies. *J. Synchrotron Radiat.* **17**, 631–643 (2010).
18. Ghiringhelli, G. *et al.* SAXES, a high resolution spectrometer for resonant x-ray emission in the 400–1600 eV energy range. *Rev. Sci. Instrum.* **77**, 113108 (2006).
19. Hozoi, L., Siurakshina, L., Fulde, P. & van den Brink, J. Ab initio determination of Cu 3d orbital energies in layered copper oxides. *Sci. Rep.* **1**, 1–4 (2011).
20. Walters, A. C. *et al.* Effect of covalent bonding on magnetism and the missing neutron intensity in copper oxide compounds. *Nature Phys.* **5**, 867–872 (2009).
21. Glawion, S. *et al.* Two-spinon and orbital excitations of the spin–Peierls system TiOCl . *Phys. Rev. Lett.* **107**, 107402 (2011).
22. Caux, J.-S. & Hagemans, R. The four-spinon dynamical structure factor of the Heisenberg chain. *J. Stat. Mech.* **2006**, P12013 (2006).
23. Oleś, A. M., Khaliullin, G., Horsch, P. & Feiner, L. F. Fingerprints of spin-orbital physics in cubic Mott insulators: magnetic exchange interactions and optical spectral weights. *Phys. Rev. B* **72**, 214431 (2005).
24. Neudert, R. *et al.* Four-band extended Hubbard Hamiltonian for the one-dimensional cuprate Sr_2CuO_3 : distribution of oxygen holes and its relation to strong intersite Coulomb interaction. *Phys. Rev. B* **62**, 10752–10765 (2000).
25. Brunner, M., Assaad, F. F. & Muramatsu, A. Single hole dynamics in the one dimensional t – J model. *Eur. Phys. J. B* **16**, 209–212 (2000).
26. Suzuura, H. & Nagaosa, N. Spin-charge separation in angle-resolved photoemission spectra. *Phys. Rev. B* **56**, 3548–3551 (1997).
27. Macfarlane, R. M. & Allen, J. W. Exciton bands in antiferromagnetic Cr_2O_3 . *Phys. Rev. B* **4**, 3054–3067 (1971).
28. Grüninger, M. *et al.* Experimental quest for orbital waves. *Nature* **418**, 39–40 (2002).
29. Ulrich, C. *et al.* Momentum dependence of orbital excitations in Mott-insulating titanates. *Phys. Rev. Lett.* **103**, 107205 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was performed at the ADRESS beamline of the Swiss Light Source using the SAXES instrument jointly built by the Paul Scherrer Institut, Switzerland, and Politecnico di Milano, Italy. We acknowledge support from the Swiss National Science Foundation and its NCCR MaNEP. K.W. acknowledges support from the Alexander von Humboldt foundation and discussions with M. Daghofer and S.-L. Drechsler. J.-S.C. acknowledges support from the Foundation for Fundamental Research on Matter and from the Netherlands Organisation for Scientific Research. S.S. and A.R. acknowledge the support of the European contract NOVMAg. This research benefited from the RIXS collaboration supported by the Computational Materials Science Network programme of the Division of Materials Science and Engineering, US Department of Energy, grant no. DE-SC0007091.

Author Contributions J.S., T.S. and H.M.R. planned the experiment. S.S. and A.R. fabricated the samples. J.S., K.J.Z., V.N.S. and T.S. carried out the experiment. J.S. and M.M. carried out the data analysis. C.M. helped with the data analysis. K.W. and J.v.d.B. developed the theory for the spin–orbital separation with assistance from M.W.H., L.H. and S.N. J.-S.C. provided the theory for the spin excitations. J.S., K.W., K.J.Z., H.M.R., J.v.d.B. and T.S. wrote the paper with contributions from all co-authors. L.P., H.M.R., J.v.d.B. and T.S. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S. (justine.schlappa@helmholtz-berlin.de) or T.S. (thorsten.schmitt@psi.ch).

METHODS

Experiment. We applied the technique of high-resolution RIXS with the incident photon energy tuned to the L_3 edge ($2p_{3/2} \rightarrow 3d$ resonance) of the copper ion (around 931 eV) and a total experimental resolution of 140 meV. The experiments were performed at the ADDRESS beamline of the Swiss Light Source at the Paul Scherrer Institut^{17,18}. Single-crystal samples of Sr_2CuO_3 were grown by the floating-zone method and freshly cleaved before the experiment. The samples were mounted such that the surface normal, [010], and the chain propagation direction, [100], were oriented parallel to the scattering plane (Supplementary Information, section 4, and Supplementary Fig. 1). The polarization vector of the incident light, \mathbf{e}_{in} , was parallel to the scattering plane (π -orientation). This yielded the maximum cross-section for the copper $2p^6 3d^9 \rightarrow 2p^5 3d^{10}$ transition. The RIXS spectrometer was located at a fixed scattering angle of either $\Psi = 90^\circ$ or 130° , collecting signal within a solid angle of $(19.3 \times 3.3) \text{ mrad}^2$ (horizontal \times vertical detection direction). The momentum transfer along the chains, q , was varied by changing the incidence angle in steps of $5 \pm 1^\circ$. This results in the RIXS signal being polarization dependent and the relative intensity of the orbital excitations varying as a function of q . (For details on the polarization dependence of the orbital excitations, see the discussion of the RIXS cross-section below.) Supplementary Fig. 2a shows the RIXS line spectra, which correspond to the intensity map data in Fig. 1c, and Supplementary Fig. 2b shows the variation in the polarization of the incident photons with q . We estimate the systematic errors in the momentum transfer to be about $0.006 \times 2\pi/a$ and $0.013 \times 2\pi/a$ at the edge and the centre of the Brillouin zone, respectively. The error in determining the energy zero of the energy transfer scale is about 30 meV. The sample was cooled with a helium-flow cryostat to 14 K during the measurements.

Theory of spin excitations. Owing to the particular choice of the experimental geometry, the RIXS cross-section for spin excitations (that is, the magnetic part of the spectrum extending up to ~ 0.8 eV) is directly proportional to the two-point dynamical spin–spin correlation function (local RIXS form factors are independent of transferred momentum here); see refs 9, 11–14. Therefore, the magnetic part of the RIXS spectrum was fitted to the two-point dynamical spin–spin correlation function (see below), convolved with a Gaussian to account for experimental resolution and adding a low-energy Gaussian covering the elastic and phonon peak around 50 meV. This captured most of the measured intensity with nearly constant J , amplitude and resolution. Fixing these three parameters to their mean values gives reasonable fits for all q . The time-normalized data sets were corrected for the effective scattering volume, V_{eff} :

$$\frac{1}{V_{\text{eff}}} = 1 + \frac{\mu_1(\omega)}{\mu_2(\omega)} \frac{\sin(\theta)}{\sin(\beta)}(q)$$

Here $(\sin(\theta)/\sin(\beta))(q)$ is a geometric term, θ and β being the angles between the sample surface and the direction of photon incidence and, respectively, the direction of photon detection; and $(\mu_1/\mu_2)(\omega)$ is an outgoing-energy-dependent absorption factor derived from the total fluorescence yield X-ray absorption spectra.

The exact solvability of the Heisenberg spin-1/2 chain is exploited to compute its two-point dynamical spin–spin correlation function. At zero magnetization, this is predominantly carried by two- and four-spinon intermediate states whose exact contributions can be written in the thermodynamic limit in terms of fundamental integrals by using the vertex operator approach²². The numerical evaluation of these integrals provides the dynamical structure factor throughout the Brillouin zone, yielding sum-rule saturations of the order of 98%.

Theory of orbital excitations. We first calculate the energies of the $3d$ orbitals in Sr_2CuO_3 using the *ab initio* quantum chemistry method (Supplementary Information, section 5). Although this method is not tailored to calculate the dispersion of orbital excitations (for which model calculations are needed) owing to the finite size of the cluster, it identifies from first principles the energies of orbital excitations in Sr_2CuO_3 for $\mathbf{q} = \mathbf{0}$ transferred momentum. These energies have been found to be in close agreement with both optical absorption and RIXS experiments¹⁹. Owing to the large separation of the d – d excitations energies (Supplementary Table 1), we can separately model the momentum-dependent RIXS cross-section for each orbital excitation for the case of non-negligible inter-orbital hopping. For this (see the next section), the Kugel'–Khomskii Hamiltonian for the case of a single orbital excitation (for example the $x^2 - y^2 \rightarrow xz$ transition) from a ferro-orbital ground state (that is, one in which all $x^2 - y^2$ orbitals are occupied) is derived on the basis of the established charge transfer model for Sr_2CuO_3 . Similar calculations are performed for the other orbital excitation symmetries (see the next section) and the RIXS cross-section based on the solution to these model calculations is calculated (see the discussion of RIXS cross-section below).

Model Hamiltonian. In Mott insulators, the Kugel'–Khomskii model is the effective low-energy superexchange model for coupled spin and orbital excitations;

compare with ref. 1. We first consider only an xz orbital excitation, which may hop between the copper sites in a 'three-step' perturbative superexchange process with an energy scale $\sim t_1 t_2 / 2U$. First the particle in the excited orbital moves from site j to the neighbouring site $j+1$ by hopping with amplitude $\sim t_2$; next an intermediate state, in which two particles are on site $j+1$ and which has an energetic cost equivalent to the on-site Coulomb repulsion $\sim U$, is formed; and finally the particle in the ground-state orbital moves from site $j+1$ to site j by hopping with amplitude t_1 . Thus, we obtain

$$H_O = -J_O \sum_{j,\sigma} (c_{j\sigma}^\dagger c_{j+1,\sigma} + \text{h.c.}) + E_O \sum_j (1 - n_j) \quad (2)$$

where $J_O = (3R_1 + R_2)2t_1 t_2 / U$ and E_O is the energy cost of the orbital excitation. In J_O , $R_1 = 1/(1 - 3\eta)$ and $R_2 = 1/(1 - \eta)$ originate in the multiplet structure of the intermediate states of the superexchange processes and depend on the ratio, $\eta = J_H/U$, of the Hund's exchange to the Coulomb repulsion. Although the above equation is given for the Mott–Hubbard limit of the superexchange model, in our calculation we modified the parameters to account for superexchange processes on oxygen atoms in Sr_2CuO_3 . For a standard set of charge transfer parameters²⁴, this gives $J_O \approx 0.075$ eV without any fitting to the experimental data. The values of E_O are determined from the RIXS experiment and are within 10% of the on-site orbital energies obtained using *ab initio* quantum chemistry cluster calculations¹⁹ for four CuO_4 plaquettes in Sr_2CuO_3 (Supplementary Information, section 5).

In equation (2), $c_{j\sigma}$ is a fermion operator, acting in the restricted Hilbert space with no double occupancies, that creates an orbital excitation (a hole in the spin background in the language of the t – J model) at site j with spin σ , and $1 - n_j = 1 - \sum_\sigma c_{j\sigma}^\dagger c_{j\sigma}$ is the number operator that counts the number of orbital excitations in the chain. Note that this 'fermionization' of the 1D problem was performed by replacing the standard orbital pseudospin operators²³ with fermions by Jordan–Wigner transformation (see ref. 8 for details).

The second part of the Hamiltonian in equation (1) corresponds to the spin dynamics on the bonds where the orbital excitation is not present:

$$H_S = J \sum_j \mathbf{S}_j \cdot \mathbf{S}_{j+1}$$

Here $J = 4t_1^2/U$ is now the well-known Heisenberg antiferromagnetic superexchange constant. In sum, H_S and H_O constitute the Hamiltonian H from equation (1).

Excitations from the ground-state $3d\ x^2 - y^2$ orbital to either the xy orbital or the xz orbital are described by H , but with different effective parameters because the hopping amplitudes from each of these orbitals to the neighbouring bonding oxygen orbitals and, subsequently, to the neighbouring copper sites differ. In particular, for the xy orbital the effective hopping amplitude along the chain is $\sim 25\%$ smaller than that for the xz orbital owing to the formation of bonding and antibonding states with $2p$ orbitals on oxygen atoms outside the copper–oxygen chain. For orbital excitations involving the yz and $3z^2 - r^2$ orbitals, the dispersion vanishes because the hopping matrix elements to the neighbouring oxygen orbitals along the chain are either much smaller than for the xy or xz orbital excitations ($3z^2 - r^2$ orbital) or are vanishingly small²⁴ (yz orbital).

RIXS cross-section. Local effective RIXS operators have been derived for the Cu^{2+} ($3d^9$) electronic configuration^{12,30} (see also Supplementary Information, section 3). However, several simplifications beyond the local effective scattering operators as used in refs 12–13, 30 are justified, because the spin of the particle in the excited orbital is to a good approximation conserved (Hund's exchange is one order of magnitude smaller than the on-site Coulomb repulsion^{23,24} during orbiton propagation). First, the spectrum obtained is independent of the spin of the particle in the excited orbital. Second, the spin and the orbital character of the excited particle are conserved during propagation. The RIXS intensity for an orbital excitation is therefore given by the sum of the intensities for scattering from initial states with spin up or down to final states with the spin either flipped or conserved. This allows us to express the RIXS intensity as the product^{11,13} of polarization-dependent intensities and the non-local dynamical structure factor $O(q, \omega)$, which is defined as the spectral function of the single orbital excitation created using operator $c_{j\sigma}$ and propagating with Hamiltonian H (ref. 8). This structure factor is either equal to one, for 'non-dispersive' orbital excitations ($3z^2 - r^2$ and yz orbitals; see previous section), or is calculated numerically by exactly solving the Hamiltonian H on a finite chain of 28 sites. The latter is done separately for xz and xy orbital excitations. We note that such a separation is possible because orbiton–orbiton interactions vanish for one-orbiton excitations in the chain. The results of these calculations are presented in Fig. 4c.

30. van Veenendaal, M. Polarization dependence of L - and M -edge resonant inelastic X-ray scattering in transition-metal compound. *Phys. Rev. Lett.* **96**, 117404 (2006).

The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis^{1,2†*}, Sohrab P. Shah^{3,4*}, Suet-Feung Chin^{1,2*}, Gulisa Turashvili^{3,4*}, Oscar M. Rueda^{1,2}, Mark J. Dunning², Doug Speed^{2,5†}, Andy G. Lynch^{1,2}, Shamith Samarajiwa^{1,2}, Yinyin Yuan^{1,2}, Stefan Gräf^{1,2}, Gavin Ha³, Gholamreza Haffari³, Ali Bashashati³, Roslin Russell², Steven McKinney^{3,4,10}, METABRIC Group[‡], Anita Langerød⁶, Andrew Green⁷, Elena Provenzano⁸, Gordon Wishart⁸, Sarah Pinder⁹, Peter Watson^{3,4,10}, Florian Markowetz^{1,2}, Leigh Murphy¹⁰, Ian Ellis⁷, Arnie Purushotham^{9,11}, Anne-Lise Børresen-Dale^{6,12}, James D. Brenton^{2,13}, Simon Tavaré^{1,2,5,14}, Carlos Caldas^{1,2,8,13} & Samuel Aparicio^{3,4}

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from representative numbers of patients. We present an integrated analysis of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in ~40% of genes, with the landscape dominated by *cis*- and *trans*-acting CNAs. By delineating expression outlier genes driven in *cis* by CNAs, we identified putative cancer genes, including deletions in *PPP2R2A*, *MTAP* and *MAP2K4*. Unsupervised analysis of paired DNA–RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, oestrogen-receptor-positive 11q13/14 *cis*-acting subgroup and a favourable prognosis subgroup devoid of CNAs. *Trans*-acting aberration hotspots were found to modulate subgroup-specific gene networks, including a TCR deletion-mediated adaptive immune response in the ‘CNA-devoid’ subgroup and a basal-specific chromosome 5 deletion-associated mitotic network. Our results provide a novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome.

Inherited genetic variation and acquired genomic aberrations contribute to breast cancer initiation and progression. Although somatically acquired CNAs are the dominant feature of sporadic breast cancers, the driver events that are selected for during tumorigenesis are difficult to elucidate as they co-occur alongside a much larger landscape of random non-pathogenic passenger alterations^{1,2} and germline copy number variants (CNVs). Attempts to define subtypes of breast cancer and to discern possible somatic drivers are still in their relative infancy^{3–6}, in part because breast cancer represents multiple diseases, implying that large numbers (many hundreds or thousands) of patients must be studied. Here we describe an integrated genomic/transcriptomic analysis of breast cancers with long-term clinical outcomes composed of a discovery set of 997 primary tumours and a validation set of 995 tumours from METABRIC (Molecular Taxonomy of Breast Cancer International Consortium).

A breast cancer population genomic resource

We assembled a collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumour banks in the UK

and Canada (Supplementary Tables 1–3). Nearly all oestrogen receptor (ER)-positive and/or lymph node (LN)-negative patients did not receive chemotherapy, whereas ER-negative and LN-positive patients did. Additionally, none of the HER2⁺ patients received trastuzumab. As such, the treatments were homogeneous with respect to clinically relevant groupings. An initial set of 997 tumours was analysed as a discovery group and a further set of 995 tumours, for which complete data later became available, was used to test the reproducibility of the integrative clusters (described below). An overview of the main analytical approaches is provided in Supplementary Fig. 1. Details concerning expression and copy number profiling, including sample assignment to the PAM50 intrinsic subtypes^{3,4,7} (Supplementary Fig. 2), copy number analysis (Supplementary Tables 4–8) and validation (Supplementary Figs 3 and 4 and Supplementary Tables 9–11), and *TP53* mutational profiling (Supplementary Fig. 5) are described in the Supplementary Information.

Genome variation affects tumour expression architecture

Genomic variants are considered to act in *cis* when a variant at a locus has an impact on its own expression, or in *trans* when it is associated

¹Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. ²Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.

³Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. ⁴Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. ⁵Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK.

⁶Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello, 0310 Oslo, Norway. ⁷Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. ⁸Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. ⁹King's College London, Breakthrough Breast Cancer Research Unit, London WC2R 2LS, UK. ¹⁰Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. ¹¹NIHR Comprehensive Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, London WC2R 2LS, UK. ¹²Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, 0316 Oslo, Norway. ¹³Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. ¹⁴Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. †Present addresses: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA (Ch.C.); University College London, Genetics Institute, WC1E 6BT, UK (D.S.).

*These authors contributed equally to this work. ‡Lists of participants and affiliations appear at the end of the paper.

©2012 Macmillan Publishers Limited. All rights reserved

with genes at other sites in the genome. We generated a map of CNAs, CNVs (Supplementary Fig. 6, Supplementary Tables 12–15) and single nucleotide polymorphisms (SNPs) in the breast cancer genome to distinguish germline from somatic variants (see Methods), and to examine the impact of each of these variants on the expression landscape. Previous studies⁸ have shown that most heritable gene expression traits are governed by a combination of *cis* (proximal) loci, defined here as those within a 3-megabase (Mb) window surrounding the gene of interest, and *trans* (distal) loci, defined here as those outside that window. We assessed the relative influence of SNPs, CNVs and CNAs on tumour expression architecture, using each of these variants as a predictor (see Methods) to elucidate expression quantitative trait loci (eQTLs) among patients.

Both germline variants and somatic aberrations were found to influence tumour expression architecture, having an impact on >39% (11,198/28,609) of expression probes genome-wide based on analysis of variance (ANOVA; see Methods), with roughly equal numbers of genes associated in *cis* and *trans*. CNAs were associated with the greatest number of expression profiles (Fig. 1, Supplementary Figs 7–13 and Supplementary Tables 16–20), but were rivalled by SNPs to explain a greater proportion of expression variation on a per-gene basis genome-wide, whereas the contribution from CNVs was more moderate (Fig. 1b and Supplementary Table 21). The true ratio of putative *trans* versus *cis* eQTLs is hard to estimate⁹; however, the large sample size used here allowed the detection of small effects, with 5,401 and 5,462 CNAs significantly (Šidák adjusted *P* value <0.0001) associated in *cis* or in *trans*, respectively. Whereas *cis*-associations tended to be stronger, the *trans*-acting loci modulated a larger number of messenger RNAs, as described below.

Expression outliers refine the breast cancer landscape

As shown above, ~20% of loci exhibit CNA-expression associations in *cis* (Supplementary Fig. 14). To refine this landscape further and identify the putative driver genes, we used profiles of outlying expression (see Methods and ref. 10) and the high resolution and sensitivity of the

Affymetrix SNP 6.0 platform to delineate candidate regions. This approach markedly reduces the complexity of the landscape to 45 regions (frequency > 5, Fig. 2) and narrows the focus, highlighting novel regions that modulate expression. The full enumeration of regions delineated by this approach and their subtype-specific associations (Supplementary Figs 15 and 16 and Supplementary Tables 22–24) includes both known drivers (for example, *ZNF703* (ref. 11), *PTEN* (ref. 12), *MYC*, *CCND1*, *MDM2*, *ERBB2*, *CCNE1* (ref. 13)) and putative driver aberrations (for example, *MDM1*, *MDM4*, *CDK3*, *CDK4*, *CAMK1D*, *PI4KB*, *NCOR1*).

The deletion landscape of breast cancer has been poorly explored, with the exception of *PTEN*. We illustrate three additional regions of significance centred on *PPP2R2A* (8p21, Fig. 2, region 11), *MTAP* (9p21, Fig. 2, region 15) and *MAP2K4* (17p11, Fig. 2, region 33), which exhibit heterozygous and homozygous deletions (Supplementary Figs 15, 17–19 and Supplementary Table 24) that drive expression of these loci. We observe breast cancer subtype-specific (enriched in mitotic ER-positive cancers) loss of transcript expression in *PPP2R2A*, a B-regulatory subunit of the PP2A mitotic exit holoenzyme complex. Somatic mutations in *PPP2R1A* have recently been reported in clear cell ovarian cancers and endometrioid cancers^{14,15}, and methylation silencing of *PPP2R2B* has also been observed in colorectal cancers¹⁶. Thus, dysregulation of specific PPP2R2A functions in luminal B breast cancers adds a significant pathophysiology to this subtype.

MTAP (9p21, a component of methyladenosine salvage) is frequently co-deleted with the *CDKN2A* and *CDKN2B* tumour suppressor genes in a variety of cancers¹⁷ as we observe here (Supplementary Figs 17c and 18). The third deletion encompasses *MAP2K4* (also called *MKK4*) (17p11), a p38/Jun dual specificity serine/threonine protein kinase. *MAP2K4* has been proposed as a recessive cancer gene¹⁸, with mutations noted in cell lines¹⁹. We show, for the first time, the recurrent deletion of *MAP2K4* (Supplementary Figs 17d and 19) concomitant with outlying expression (Supplementary Fig. 15) in predominantly ER-positive cases, and verify homozygous deletions (Supplementary Table 9) in primary tumours, strengthening the evidence for *MAP2K4* as a tumour suppressor in breast cancer.

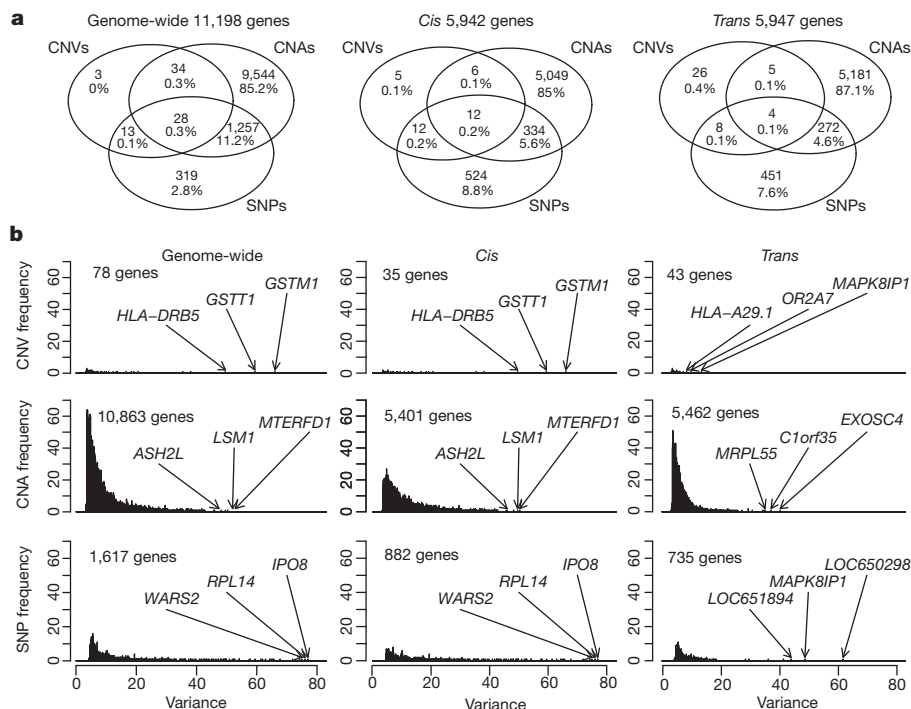


Figure 1 | Germline and somatic variants influence tumour expression architecture. **a**, Venn diagrams depict the relative contribution of SNPs, CNVs and CNAs to genome-wide, *cis* and *trans* tumour expression variation for the most significant expression associations (Šidák adjusted *P*-value ≤ 0.0001).

b, Histograms illustrate the proportion of variance explained by the most significantly associated predictor type for each predictor type, where several of the top associations are indicated.

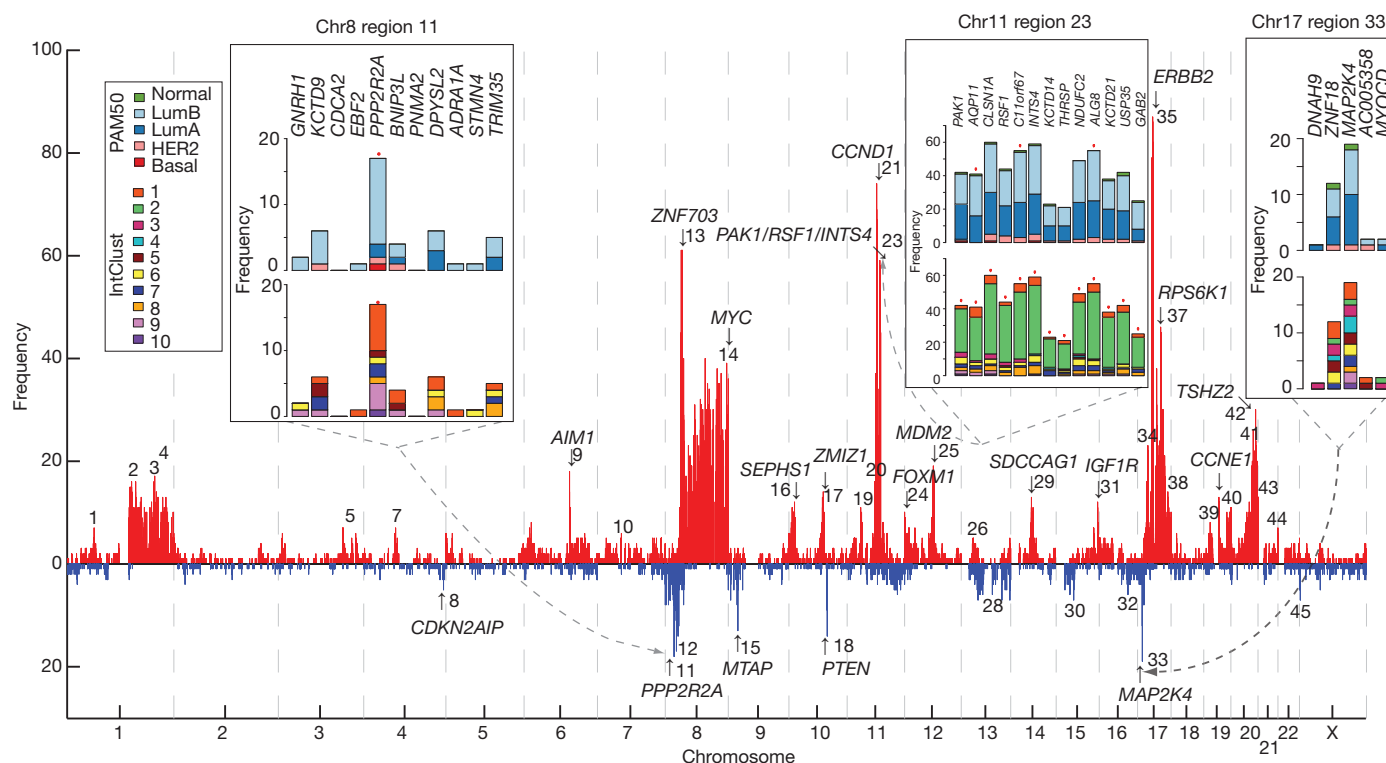


Figure 2 | Patterns of *cis* outlying expression refine putative breast cancer drivers. A genome-wide view of outlying expression coincident with extreme copy number events in the CNA landscape highlights putative driver genes, as indicated by the arrows and numbered regions. The frequency (absolute count) of cases exhibiting an outlying expression profile at regions across the genome is

shown, as is the distribution across subgroups for several regions in the insets. High-level amplifications are indicated in red and homozygous deletions in blue. Red asterisks above the bar plots indicate significantly different observed distributions than expected based on the overall population frequency (χ^2 test, $P < 0.0001$).

Trans-acting associations reveal distinct modules

We next asked how *trans*-associated expression profiles are distributed across the genome. We mapped these in the expression landscape by examining the matrices of CNA–expression associations (see Methods). This revealed strong off-diagonal patterns at loci on chromosomes 1q, 7p, 8, 11q, 14q, 16, 17q and 20q (Fig. 3a), including both positive and negative associations, as well as numerous *trans*-acting aberration hotspots (defined as CNAs associated with >30 mRNAs). Importantly, these aberration hotspots can be grouped into pathway modules, which highlight known driver loci such as *ERBB2* and *MYC*, as well as novel loci associated with large *trans* expression modules (Supplementary Tables 25 and 26). The T-cell-receptor (TCR) loci on chromosomes 7 (*TRG*) and 14 (*TRA*) represent two such hotspots that modulated 381 and 153 unique mRNAs, respectively, as well as 19 dually regulated genes (Supplementary Fig. 20). These cognate mRNAs were highly enriched for T-cell activation and proliferation, dendritic cell presentation, and leukocyte activation, which indicate the induction of an adaptive immune response associated with tumour-infiltrating lymphocytes (Fig. 3b, Supplementary Fig. 20 and Supplementary Tables 27 and 28), as described later.

In a second approach, we examined the genome-wide patterns of linear correlation between copy number and expression features (see Methods), and noted the alignment of several off-diagonal signals, including those on chromosome 1q, 8q, 11q, 14q and 16 (Supplementary Fig. 21). Additionally, a broad signal on chromosome 5 localizing to a deletion event restricted to the basal-like tumours was observed (Supplementary Fig. 21), but was not detected with the eQTL framework, where discrete (as opposed to continuous) copy number values were used. This basal-specific *trans* module is enriched for transcriptional changes involving cell cycle, DNA damage repair and apoptosis (Supplementary Table 29), reflecting the high mitotic index typically associated with basal-like tumours, described in detail below.

Integrative clustering reveals novel subgroups

Using the discovery set of 997 breast cancers, we next asked whether novel biological subgroups could be found by joint clustering of copy number and gene expression data. On the basis of our finding that *cis*-acting CNAs dominated the expression landscape, the top 1,000 *cis*-associated genes across all subtypes (Supplementary Table 30) were used as features for input to a joint latent variable framework for integrative clustering²⁰ (see Methods). Cluster analysis suggested 10 groups (based on Dunn's index) (see Methods and Supplementary Figs 22 and 23), but for completeness, this result was compared with the results for alternative numbers of clusters and clustering schemes (see Methods, Supplementary Figs 23–27 and Supplementary Tables 31–33). The 10 integrative clusters (labelled IntClust 1–10) were typified by well-defined copy number aberrations (Fig. 4, Supplementary Figs 22, 28–30 and Supplementary Tables 34–39), and split many of the intrinsic subtypes (Supplementary Figs 31–33). Kaplan–Meier plots of disease-specific survival and Cox proportional hazards models indicate subgroups with distinct clinical outcomes (Fig. 5, Supplementary Figs 34, 35 and Supplementary Tables 40 and 41). To validate these results, we trained a classifier (754 features) for the integrative subtypes in the discovery set using the nearest shrunken centroids approach²¹ (see Methods and Supplementary Tables 42 and 43), and then classified the independent validation set of 995 cases into the 10 groups (Supplementary Table 44). The reproducibility of the clusters in the validation set is shown in three ways. First, classification of the validation set resulted in the assignment of a similar proportion of cases to the 10 subgroups, each of which exhibited nearly identical copy number profiles (Fig. 4). Second, the groups have substantially similar hazard ratios (Fig. 5b, Supplementary Fig. 35 and Supplementary Table 40). Third, the quality of the clusters in the validation set is emphasized by the in-group proportions (IGP) measure²² (Fig. 4).

Among the integrative clusters, we first note an ER-positive subgroup composed of 11q13/14 *cis*-acting luminal tumours (IntClust 2,

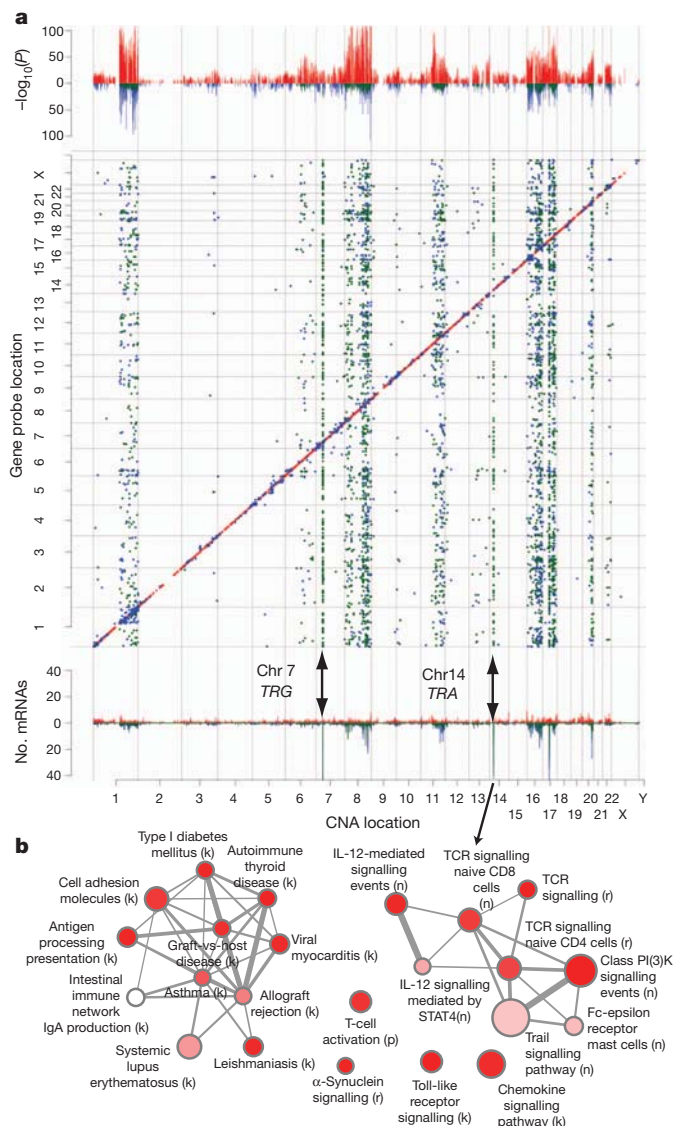


Figure 3 | Trans-acting aberration hotspots modulate concerted molecular pathways. **a**, Manhattan plot illustrating *cis* and *trans* expression-associated copy number aberrations from the eQTL analysis (top panel). The matrix of significant predictor–expression associations (adjusted P -value ≤ 0.0001) exhibits strong off-diagonal patterns (middle panel), and the frequency of mRNAs associated with a particular copy number aberration further illuminates these *trans*-acting aberration hotspots (bottom panel). The directionality of the associations is indicated as follows: *cis*: positive, red; negative, pink; *trans*: positive, blue; negative, green. **b**, Enrichment map of immune response modules in the *trans*-associated TRA network, where letters in parentheses represent the source database as follows: b, NCI-PID BioCarta; c, cancer cell map; k, KEGG; n, NCI-PID curated pathways; p, PANTHER; r, Reactome.

$n = 45$) that harbour other common alterations. This subgroup exhibited a steep mortality trajectory with elevated hazard ratios (discovery set: 3.620, 95% confidence interval (1.905–6.878); validation set: 3.353, 95% confidence interval (1.381–8.141)), indicating that it represents a particularly high-risk subgroup. Several known and putative driver genes reside in this region, namely *CCND1* (11q13.3), *EMSY* (11q13.5), *PAK1* (11q14.1) and *RSF1* (11q14.1), which have been previously linked to breast^{13,23} or ovarian cancer²⁴. Both the copy number (Fig. 4) and expression outlier landscapes (Fig. 2) suggest at least two separate amplicons at 11q13/14, one at *CCND1* (11q13.3) and a separate peak from 11q13.5–11q14.1 spanning *UVRAG*–*GAB2*, centred around *PAK1*, *RSF1*, *C11orf67* and *INTS4*, where it is more challenging to distinguish the driver²⁴. Notably, the

expression outlier profiles for this region are enriched for samples belonging to IntClust 2 (Fig. 2, inset region 23) and all 45 members of this subgroup harboured amplifications of these genes, with high frequencies of amplification also observed for *CCND1* ($n = 39$) and *EMSY* ($n = 34$). In light of these observations, the 11q13/14 amplicon may be driven by a cassette of genes rather than a single oncogene.

Second, we note the existence of two subgroups marked by a paucity of copy number and *cis*-acting alterations. These subgroups cannot be explained by low cellularity tumours (see Methods). One subgroup (IntClust3, $n = 156$) with low genomic instability (Fig. 4 and Supplementary Fig. 22) was composed predominantly of luminal A cases, and was enriched for histotypes that typically have good prognosis, including invasive lobular and tubular carcinomas. The other subgroup (IntClust 4, $n = 167$) was also composed of favourable outcome cases, but included both ER-positive and ER-negative cases and varied intrinsic subtypes, and had an essentially flat copy number landscape, hence termed the ‘CNA-devoid’ subgroup. A significant proportion of cases within this subgroup exhibit extensive lymphocytic infiltration (Supplementary Table 45).

Third, several intermediate prognosis groups of predominantly ER-positive cancers were identified, including a 17q23/20q *cis*-acting luminal B subgroup (IntClust 1, $n = 76$), an 8p12 *cis*-acting luminal subgroup (IntClust 6, $n = 44$), as well as an 8q *cis*-acting/20q-amplified mixed subgroup (IntClust 9, $n = 67$). Two luminal A subgroups with similar CNA profiles and favourable outcome were noted. One subgroup is characterized by the classical 1q gain/16q loss (IntClust 8, $n = 143$), which corresponds to a common translocation event²⁵, and the other lacks the 1q alteration, while maintaining the 16p gain/16q loss with higher frequencies of 8q amplification (IntClust 7, $n = 109$). We also noted that the majority of basal-like tumours formed a stable, mostly high-genomic instability subgroup (IntClust 10, $n = 96$). This subgroup had relatively good long-term outcomes (after 5 years), consistent with ref. 26, and characteristic *cis*-acting alterations (5 loss/8q gain/10p gain/12p gain).

The *ERBB2*-amplified cancers composed of HER2-enriched (ER-negative) cases and luminal (ER-positive) cases appear as IntClust 5 ($n = 94$), thus refining the *ERBB2* intrinsic subtype by grouping additional patients that might benefit from targeted therapy. Patients in this study were enrolled before the general availability of trastuzumab, and as expected this subgroup exhibits the worst disease-specific survival at both 5 and 15 years and elevated hazard ratios (discovery set: 3.899, 95% confidence interval (2.234–6.804); validation set: 4.447, 95% confidence interval (2.284–8.661)).

Pathway deregulation in the integrative subgroups

Finally, we projected the molecular profiles of the integrative subgroups onto pathways to examine possible biological themes among breast cancer subgroups (Supplementary Tables 46 and 47) and the relative impact of *cis* and *trans* expression modules on the pathways. The CNA-devoid (IntClust 4) group exhibits a strong immune and inflammation signature involving the antigen presentation pathway, OX40 signalling, and cytotoxic T-lymphocyte-mediated apoptosis (Supplementary Fig. 36). Given that *trans*-acting deletion hotspots were localized to the *TRG* and *TRA* loci and were associated with an adaptive immune response module, we asked whether these deletions contribute to alterations in this pathway. The CNA-devoid subgroup (IntClust 4) was found to exhibit nearly twice as many deletions (typically heterozygous loss) at the *TRG* and *TRA* loci ($\sim 20\%$ of cases) as compared to the other subtypes (with the exception of IntClust 10), and deletions of both TCR loci were significantly associated with severe lymphocytic infiltration (χ^2 test, $P < 10^{-9}$ and $P < 10^{-8}$, respectively). Notably, these *trans*-associated mRNAs were significantly enriched in the immune response signature of the CNA-devoid subgroup (Supplementary Fig. 36) as well as among genes differentially expressed in CNA-devoid cases with severe lymphocytic infiltration (Supplementary Fig. 37). We conclude that genomic copy number loss

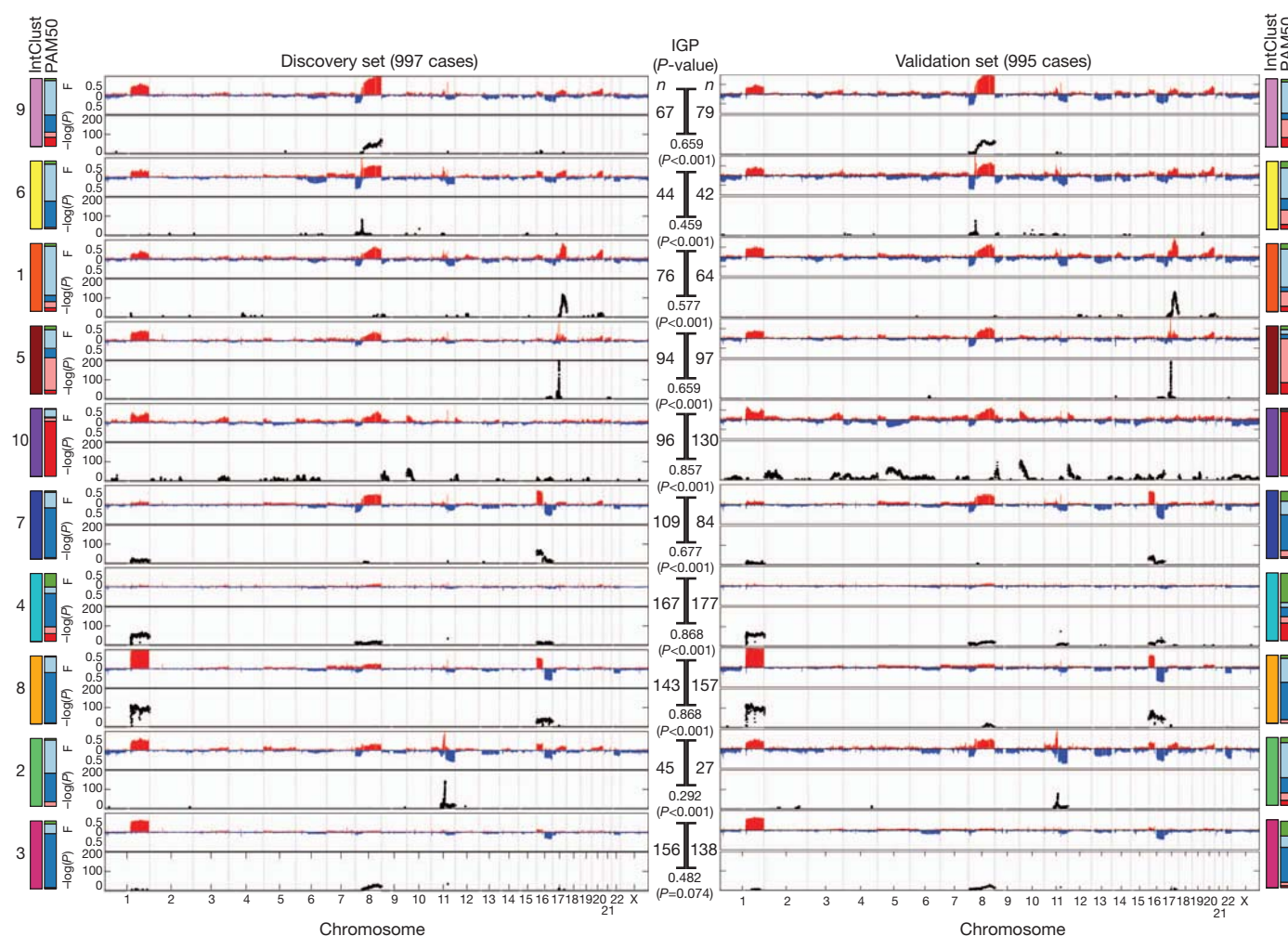


Figure 4 | The integrative subgroups have distinct copy number profiles. Genome-wide frequencies (F, proportion of cases) of somatic CNAs (y-axis, upper plot) and the subtype-specific association ($-\log_{10} P$ -value) of aberrations (y-axis, bottom plot) based on a χ^2 test of independence are shown for each of the 10 integrative clusters. Regions of copy number gain are indicated in red and regions of loss in blue in the frequency plot (upper plot). Subgroups were

ordered by hierarchical clustering of their copy number profiles in the discovery cohort ($n = 997$). For the validation cohort ($n = 995$), samples were classified into each of the integrative clusters as described in the text. The number of cases in each subgroup (n) is indicated as is the in-group proportion (IGP) and associated P -value, as well as the distribution of PAM50 subtypes within each cluster.

at the TCR loci drives a *trans*-acting immune response module that associates with lymphocytic infiltration, and characterizes an otherwise genomically quiescent subgroup of ER-positive and ER-negative patients with good prognosis. These observations suggest the presence of mature T lymphocytes (with rearranged TCR loci), which may explain an immunological response to the cancer. In line with these findings, a recent study²⁷ demonstrated the association between CD8⁺ lymphocytes and favourable prognosis.

Also among the *trans*-influenced groups is IntClust 10 (basal-like cancer enriched subgroup), which harbours chromosome 5q deletions (Supplementary Fig. 21). Numerous signalling molecules, transcription factors and cell division genes were associated in *trans* with this deletion event in the basal cancers, including alterations in *AURKB*, *BCL2*, *BUB1*, *CDCA3*, *CDCA4*, *CDC20*, *CDC45*, *CHEK1*, *FOXM1*, *HDAC2*, *IGF1R*, *KIF2C*, *KIFC1*, *MTHFD1L*, *RAD51AP1*, *TTK* and *UBE2C* (Supplementary Fig. 38). Notably, *TTK* (*MPS1*), a dual specificity kinase that assists *AURKB* in chromosome alignment during mitosis, and recently reported to promote aneuploidy in breast cancer²⁸, was upregulated. These results indicate that 5q deletions modulate the coordinate transcriptional control of genomic and chromosomal instability and cell cycle regulation within this subgroup.

In contrast to these subtype-specific *trans*-associated signatures, the high-risk 11q13/14 subgroup was characterized by strong

cis-acting associations. Like the basal cancers, this subgroup also exhibited alterations in key cell-cycle-related genes (Supplementary Fig. 39), which probably have a role in its aggressive pathophysiology, but the nature of the signature differs. In particular, the regulation of the G1/S transition by BTG family proteins, which include *CCND1*, *PPP2R1B* and *E2F2*, was significantly enriched in the 11q13/14 *cis*-acting subgroup, but not the basal cancers, and this is consistent with *CCND1* and the *PPP2R* subunit representing subtype-specific drivers in these tumours.

Discussion

We have generated a robust, population-based molecular subgrouping of breast cancer based on multiple genomic views. The size and nature of this cohort made it amenable to eQTL analyses, which can aid the identification of loci that contribute to the disease phenotype²⁹. CNAs and SNPs influenced expression variation, with CNAs dominating the landscape in *cis* and *trans*. The joint clustering of CNAs and gene expression profiles further resolves the considerable heterogeneity of the expression-only subgroups, and highlights a high-risk 11q13/14 *cis*-acting subgroup as well as several other strong *cis*-acting clusters and a genomically quiescent group. The reproducibility of subgroups with these molecular and clinical features in a validation cohort of 995 tumours suggests that by integrating multiple genomic

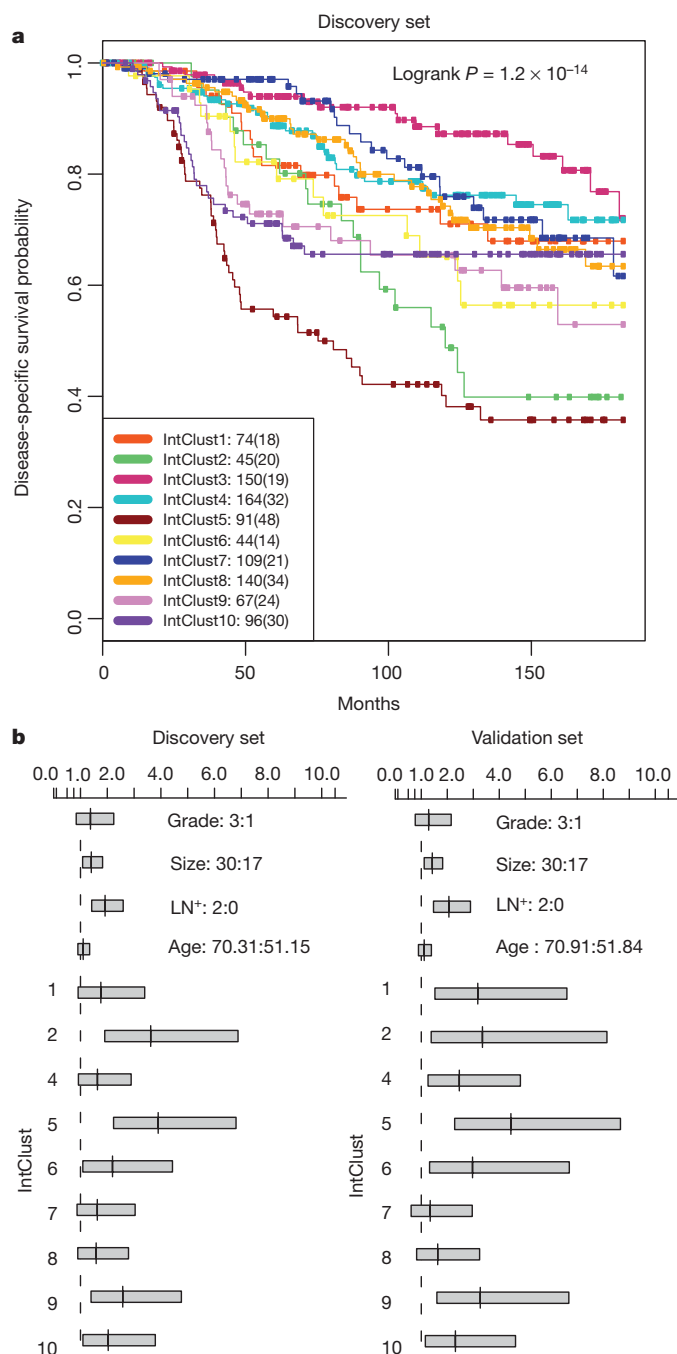


Figure 5 | The integrative subgroups have distinct clinical outcomes.

a, Kaplan-Meier plot of disease-specific survival (truncated at 15 years) for the integrative subgroups in the discovery cohort. For each cluster, the number of samples at risk is indicated as well as the total number of deaths (in parentheses). **b**, 95% confidence intervals for the Cox proportional hazard ratios are illustrated for the discovery and validation cohort for selected values of key covariates, where each subgroup was compared against IntClust 3.

features it may be possible to derive more robust patient classifiers. We show here, for the first time, that subtype-specific *trans*-acting aberrations modulate concerted transcriptional changes, such as the TCR deletion-mediated adaptive immune response that characterizes the CNA-devoid subgroup and the chromosome 5 deletion-associated cell cycle program in the basal cancers.

The integrated CNA-expression landscape highlights a limited number of genomic regions that probably contain driver genes, including *ZNF703*, which we recently described as a luminal B specific driver¹¹, as well as somatic deletion events affecting key subunits of the

PP2A holoenzyme complex and *MTAP*, which have previously been under-explored in breast cancer. The CNA-expression landscape also illuminates rare but potentially significant events, including *IGF1R*, *KRAS* and *EGFR* amplifications and *CDKN2B*, *BRCA2*, *RBI*, *ATM*, *SMAD4*, *NCOR1* and *UTX* homozygous deletions. Although some of these events have low overall frequencies (<1% patients) (Figs 2, Supplementary Fig. 15 and Supplementary Tables 22–24), they may have implications for understanding therapeutic responses to targeted agents, particularly those targeting tyrosine kinases or phosphatases.

Finally, because the integrative subgroups occur at different frequencies in the overall population, focusing sequencing efforts on representative numbers from these groups will help to establish a comprehensive breast cancer somatic landscape at sequence-level resolution. For example, a significant number (~17%, $n = 167$ in the discovery cohort) of breast cancers are devoid of somatic CNAs, and are ripe for mutational profiling. Our work provides a definitive framework for understanding how gene copy number aberrations affect gene expression in breast cancer and reveals novel subgroups that should be the target of future investigation.

METHODS SUMMARY

All patient specimens were obtained with appropriate consent from the relevant institutional review board. DNA and RNA were isolated from samples and hybridized to the Affymetrix SNP 6.0 and Illumina HT-12 v3 platforms for genomic and transcriptional profiling, respectively. A detailed description of the experimental assays and analytical methods used to analyse these data are available in the Supplementary Information.

Received 24 April 2011; accepted 22 February 2012.

Published online 18 April 2012.

- Leary, R. J. *et al.* Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 16224–16229 (2008).
- Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Sørli, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
- Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529–541 (2006).
- Chin, S. F. *et al.* High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8**, R215 (2007).
- Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78 (2005).
- Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).
- Teschendorff, A. E., Naderi, A., Barbosa-Morais, N. L. & Caldas, C. PACK: Profile analysis using clustering and kurtosis to find molecular classifiers in cancer. *Bioinformatics* **22**, 2269–2275 (2006).
- Holland, D. *et al.* *ZNF703* is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol. Med.* **3**, 167–180 (2011).
- Li, J. *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943–1947 (1997).
- Santarius, T., Shiply, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nature Rev. Cancer* **10**, 59–64 (2010).
- Jones, S. *et al.* Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
- McConechy, M. K. *et al.* Subtype-specific mutation of *PPP2R1A* in endometrial and ovarian carcinomas. *J. Pathol.* **223**, 567–573 (2011).
- Tan, J. *et al.* B55 β -associated PP2A complex controls PDK1-directed MYC signaling and modulates rapamycin sensitivity in colorectal cancer. *Cancer Cell* **18**, 459–471 (2010).
- Christopher, S. A., Diegelman, P., Porter, C. W. & Kruger, W. D. Methylthioadenosine phosphorylase, a gene frequently codeleted with p16 (*CDKN2A/ARF*), acts as a tumor suppressor in a breast cancer cell line. *Cancer Res.* **62**, 6639–6644 (2002).
- Teng, D. H. *et al.* Human mitogen-activated protein kinase kinase 4 as a candidate tumor suppressor. *Cancer Res.* **57**, 4177–4182 (1997).
- Holstelt, A. *et al.* Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **121**, 53–64 (2010).

20. Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906–2912 (2009).
21. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunk centroids of gene expression. *Proc. Natl Acad. Sci. USA* **99**, 6567–6572 (2002).
22. Kapp, A. V. & Tibshirani, R. Are clusters found in one dataset present in another dataset? *Biostatistics* **8**, 9–31 (2007).
23. Hughes-Davies, L. *et al.* EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* **115**, 523–535 (2003).
24. Brown, L. A. *et al.* Amplification of 11q13 in ovarian carcinoma. *Genes Chromosom. Cancer* **47**, 481–489 (2008).
25. Russnes, H. G. *et al.* Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl. Med.* **2**, 38ra47 (2010).
26. Blows, F. M. *et al.* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7**, e1000279 (2010).
27. Mahmoud, S. M. A. *et al.* Tumor-infiltrating CD8⁺ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* **29**, 1949–1955 (2011).
28. Daniel, J., Coulter, J., Woo, J.-H., Wilsbach, K. & Gabrielson, E. High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. *Proc. Natl Acad. Sci. USA* **108**, 5384–5389 (2011).
29. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The METABRIC project was funded by Cancer Research UK, the British Columbia Cancer Foundation and Canadian Breast Cancer Foundation BC/Yukon. The authors also acknowledge the support of the University of Cambridge, Hutchinson Whampoa, the NIHR Cambridge Biomedical Research Centre, the Cambridge Experimental Cancer Medicine Centre, the Centre for Translational Genomics (CTAG) Vancouver and the BCCA Breast Cancer Outcomes Unit. S.P.S. is a Michael Smith Foundation for Health Research fellow. S.A. is supported by a Canada Research Chair. This work was supported by the National Institutes of Health Centers of Excellence in Genomics Science grant P50 HG02790 (S.T.). The authors thank C. Perou and J. Parker for discussions on the use of the PAM50 centroids. They also acknowledge the patients who donated tissue and the associated pseudo-anonymized clinical data for this project.

Author Contributions Ch.C. led the analysis, designed experiments and wrote the manuscript. S.P.S. led the HMM-based analyses, expression outlier and *TP53* analyses, and contributed to manuscript preparation. S.-F.C. generated data, designed and performed experiments. G.T. generated data, provided histopathology expertise and analysed *TP53* sequence data. O.M.R., M.J.D., D.S., A.G.L., S.S., Y.Y., S.G., Ga.H., Gh.H., A.B., R.R., S.M. and F.M. performed analyses. G.T., A.G., E.P., S.P. and I.E. provided histopathology expertise. A.L. performed *TP53* sequencing. A.-L.B.-D. oversaw *TP53* sequencing. S.P., P.W., L.M., G.W., I.E., A.P., Ca.C. and S.A. contributed to sample selection. J.D.B. and S.T. contributed to study design. S.T. provided statistical expertise. The METABRIC Group contributed collectively to this study. Ca.C. and S.A. co-conceived and oversaw the study, and contributed to manuscript preparation and were responsible for final editing. Ca.C. and S.A. are joint senior authors and project co-leaders.

Author Information The associated genotype and expression data have been deposited at the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>), which is hosted by the European Bioinformatics Institute, under accession number EGAS00000000083. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Ca.C. (carlos.caldas@cancer.org.uk) or S.A. (saparicio@bccrc.ca).

METABRIC Group

Co-chairs Carlos Caldas^{1,2}, Samuel Aparicio^{3,4}

Writing committee Christina Curtis^{1,2†}, Sohrab P. Shah^{3,4}, Carlos Caldas^{1,2}, Samuel Aparicio^{3,4}

Steering committee James D. Brenton^{1,2}, Ian Ellis⁵, David Huntsman^{3,4}, Sarah Pinder⁶, Arnie Purushotham⁶, Leigh Murphy⁷, Carlos Caldas^{1,2}, Samuel Aparicio^{3,4}

Tissue and clinical data source sites: **University of Cambridge/Cancer Research UK Cambridge Research Institute** Carlos Caldas (Principal Investigator)^{1,2}; Helen Bardwell², Suet-Feung Chin^{1,2}, Christina Curtis^{1,2†}, Zhihao Ding², Stefan Gräf^{1,2}, Linda Jones⁸, Bin Liu^{1,2}, Andy G. Lynch^{1,2}, Irene Papatheodorou^{1,2}, Stephen J. Sammut⁹, Gordon Wishart⁹; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)^{3,4}, Steven Chia⁴, Karen Gelmon⁴, David Huntsman^{3,4}, Steven McKinney^{3,4}, Caroline Speers⁴, Gulisa Turashvili^{3,4}, Peter Watson^{3,4,7}; **University of Nottingham:** Ian Ellis (Principal Investigator)⁵, Roger Blamey⁵, Andrew Green⁵, Douglas Macmillan⁵, Emad Rakha⁵; **King's College London** Arnie Purushotham (Principal Investigator)⁶, Cheryl Gillett⁶, Anita Grigoriadis⁶, Sarah Pinder⁶, Emanuele di Rinaldis⁶, Andy Tutt⁶; **Manitoba Institute of Cell Biology** Leigh Murphy (Principal Investigator)⁷, Michelle Parisien⁷, Sandra Troup⁷

Cancer genome/transcriptome characterization centres: **University of Cambridge/Cancer Research UK Cambridge Research Institute** Carlos Caldas (Principal Investigator)^{1,2}, Suet-Feung Chin (Team Leader)^{1,2}, Derek Chan¹, Claire Fielding², Ana-Teresa Maia^{1,2}, Sarah McGuire², Michelle Osborne², Sara M. Sayalero², Inmaculada Spiteri², James Hadfield²; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)^{3,4}, Gulisa Turashvili (Team Leader)^{3,4}, Lynda Bell⁴, Katie Chow⁴, Nadia Gale⁴, David Huntsman^{3,4}, Maria Kovalik⁴, Ying Ng⁴, Leah Prentice⁴

Data analysis subgroup: **University of Cambridge/Cancer Research UK Cambridge Research Institute** Carlos Caldas (Principal Investigator)^{1,2}, Simon Tavaré (Principal Investigator)^{1,2,10,11}, Christina Curtis (Team Leader)^{1,2†}, Mark J. Dunning², Stefan Gräf^{1,2}, Andy G. Lynch^{1,2}, Oscar M. Rueda^{1,2}, Roslin Russell², Shamith Samarajiva^{1,2}, Doug Speed^{2,10}, Florian Markowetz (Principal Investigator)^{1,2}, Yinyin Yuan^{1,2}; James D. Brenton (Principal Investigator)^{1,2}; **British Columbia Cancer Agency** Samuel Aparicio (Principal Investigator)^{3,4}, Sohrab P. Shah (Team Leader)^{3,4}, Ali Bashashati³, Gavin Ha³, Gholamreza Haffari³ & Steven McKinney^{3,4}

¹Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. ²Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ³Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. ⁴Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V5Z 1L3, Canada. ⁵Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. ⁶King's College London, Breakthrough Breast Cancer Research Unit, London, WC2R 2LS, UK. ⁷Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada. ⁸Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. ⁹Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. ¹⁰Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. ¹¹Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. †Present address: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.

Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding

Eliaana Marinari^{1*}, Aida Mehonic^{2*}, Scott Curran¹, Jonathan Gale³, Thomas Duke² & Buzz Baum¹

The development and maintenance of an epithelium requires finely balanced rates of growth and cell death. However, the mechanical and biochemical mechanisms that ensure proper feedback control of tissue growth^{1–4}, which when deregulated contribute to tumorigenesis, are poorly understood. Here we use the fly notum as a model system⁵ to identify a novel process of crowding-induced cell delamination that balances growth to ensure the development of well-ordered cell packing. In crowded regions of the tissue, a proportion of cells undergo a serial loss of cell–cell junctions and a progressive loss of apical area, before being squeezed out by their neighbours. This path of delamination is recapitulated by a simple computational model of epithelial mechanics, in which stochastic cell loss relieves overcrowding as the system tends towards equilibrium. We show that this process of delamination is mechanistically distinct from apoptosis-mediated cell extrusion^{6–8} and precedes the first signs of cell death. Overall, this analysis reveals a simple mechanism that buffers epithelia against variations in growth. Because live-cell delamination constitutes a mechanistic link between epithelial hyperplasia and cell invasion, this is likely to have important implications for our understanding of the early stages of cancer development.

To better understand the role of epithelial cell dynamics and mechanics in the development of a well-ordered tissue, we studied the refinement of the fly notum^{5,9} from 12 h after pupariation (AP), as the tissue undergoes a final process of topological reorganization before terminal differentiation. The refinement of cell packing in this tissue occurs largely without concomitant changes in overall tissue form or size^{5,9}. Using the constitutive expression of E-cadherin–GFP to follow apical junctional dynamics (Supplementary Movie 1), we observed a significant level of regionalized epithelial cell delamination in the tissue. In the midline region, where cells have an anisotropic geometry at 12 h AP, $32.5 \pm 13.6\%$ of cells underwent basal delamination, compared with $0.5 \pm 0.3\%$ of cells in the surrounding tissue (means \pm s.d., $n = 4$ pupae) (Fig. 1c). The pattern of delamination varied between animals, was not symmetric across the midline, and involved a variable proportion of cells leaving the tissue both before and after cell division (Fig. 1b). This rules out a deterministic role for cell lineage, position or developmental time in delamination, implying that it is a stochastic process. Before the onset of delamination, cells within the midline of the tissue were 2.7-fold longer than they were wide when measured along the anterior–posterior axis of the animal, despite their having similar apical areas ($48.5 \pm 14.2 \mu\text{m}^2$ in the midline and $51.3 \pm 12.7 \mu\text{m}^2$ outside the midline). By 26 h AP, however, midline cells were isotropic in form and could no longer be distinguished from those in the rest of the tissue (Fig. 1a). Regionalized delamination therefore accompanies the topological rearrangement as the tissue tends towards a uniform cell organization. This suggests a possible role for delamination in the refinement of this tissue before terminal differentiation.

Although it is not currently known how epithelial growth is regulated during development or homeostasis, mechanical stresses generated as

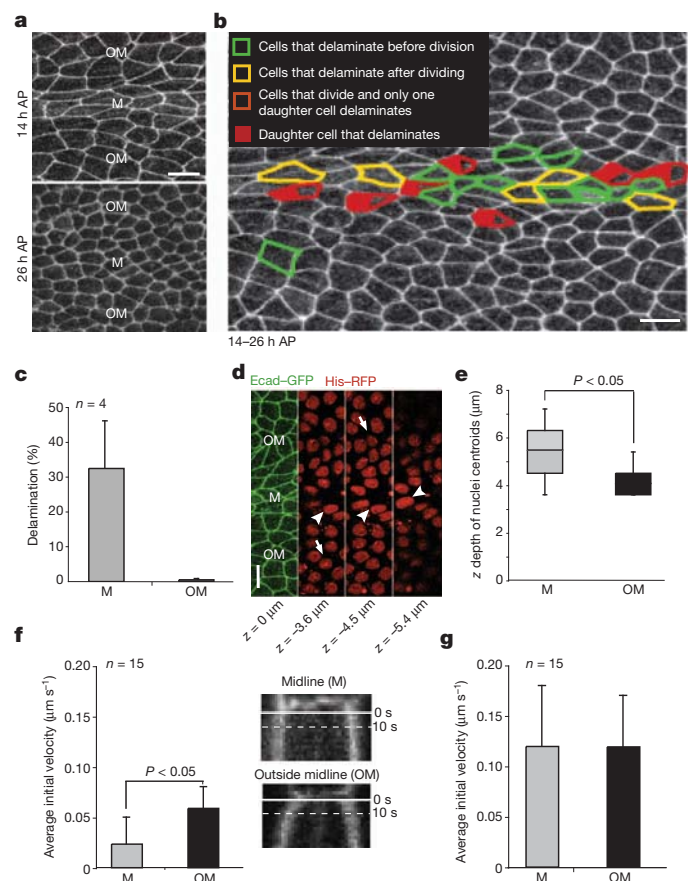


Figure 1 | Stochastic cell delamination from crowded regions of a tissue during normal development. **a**, E-cadherin–GFP notum at 14 h and 26 h AP. **b**, Cells delaminating between 14 h and 26 h AP were coloured according to their behaviour (see key). **c**, The percentage of cells delaminating in the midline (M) and outside the midline (OM), calculated as the ratio between the number of cells that delaminated in 11 h and the total initial number of cells from four animals. **d**, Tissue expressing E-cadherin–GFP and His–RFP (to mark nuclei) along different planes in *z*. Midline nuclei are indicated by arrowheads and those outside the midline by arrows. **e**, Box plot showing the *z*-depth of nuclei centroids in M and OM. The median is represented by horizontal lines; the 75th and 25th centiles are at the top and bottom of the boxes, respectively. **f**, Quantification and kymograph analysis of average initial velocity of vertex displacement after laser cutting for early pupae (14–16 h AP). Images show the junction before the cut at 0 s and the kymograph after the cut. Dotted lines indicate 10 s after cutting. The average initial velocity does not correlate with junction length (Supplementary Fig. 2b–d). **g**, Quantification of the average initial velocity of vertex displacement of single junctions after laser ablation for late pupae (24–26 h AP). In **f** and **g**, *n* represents 15 cuts for 5 animals. All error bars indicate \pm s.d. Scale bars, 10 μm .

¹Medical Research Council Laboratory of Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, UK. ²London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK. ³UCL Ear Institute, University College London, 332 Gray's Inn Road, London WC1X 8EE, UK.

*These authors contributed equally to this work.

a consequence of cell proliferation have been proposed to have a function^{1,4}. Several lines of evidence suggest that cells across the developing notum are subject to regional differences in forces that could, through mechanical feedback, lead to local differences in cell behaviour. First, at 14 h AP, nuclei in the midline region are found at variable depths along the *z* axis (Fig. 1d,e), whereas nuclei outside the midline lie in the same confocal plane (Fig. 1d,e). At the same time, cell–cell junctions within the midline region seem convoluted, whereas those outside the midline seem taut (Supplementary Fig. 1a). Furthermore, when junctions are cut with a diffraction limited laser, there is a threefold difference in the initial velocity of the junctional recoil in these two regions. Thus, the vertices of severed junctions outside the midline open with an initial velocity of $0.06 \pm 0.02 \mu\text{m s}^{-1}$, whereas vertices in the midline remain relatively unmoved ($0.02 \pm 0.03 \mu\text{m s}^{-1}$; Fig. 1f and Supplementary Fig. 2a). In contrast, by 26 h AP, when nuclear and junctional morphology seem homogeneous across the tissue (Supplementary Fig. 1b), rates of vertex displacement are similar within the midline ($0.12 \pm 0.06 \mu\text{m s}^{-1}$) and outside it ($0.12 \pm 0.05 \mu\text{m s}^{-1}$; Fig. 1g and Supplementary Fig. 2a). These differences in tissue mechanics were confirmed by using different patterns of laser-induced dissection (Supplementary Figs 3 and 4). Because local delamination rates correlate with local differences in mechanics across the tissue, delamination could relieve local overcrowding to help generate the regular cell packing and uniform tissue mechanics seen at the end of development.

To test this hypothesis, we used RNA-mediated interference (RNAi) to determine whether rates of cell delamination are sensitive to changes in cell crowding. We began by silencing components of the phosphatidylinositol-3-OH kinase (PI(3)K) and Hippo pathways (Fig. 2a,b and Supplementary Fig. 5). These treatments increased cell growth without significantly increasing tissue area, resulting in an increase in epithelial height and a corresponding increase in the rate of tissue-wide cell delamination (Fig. 2a–d and Supplementary Movie 2). Rates of epithelial cell delamination were highest in the midline and in crowded epithelial folds that formed as the result of tissue buckling (Fig. 2g). Conversely, rates of midline cell delamination were significantly decreased when growth and crowding were inhibited (Fig. 2a–d,f and Supplementary Movies 3 and 4). Moreover, in tissue expressing high levels of Tsc1 and Tsc2 the decrease in growth was accompanied by a corresponding increase in junctional tension across the tissue, measured by vertex displacement after cutting (Fig. 2e), as expected if crowding affects tissue mechanics. This marked decrease in the rate of delamination seen in tissues with decreased growth was not a simple consequence of the accompanying decrease in cell division, because delamination continued unabated when cell division was directly inhibited through the depletion of Cdc25/String (Supplementary Fig. 6 and Supplementary Movie 5). Furthermore, differences in the rate of delamination could not be explained by growth-induced changes in developmental timing, because similar results were seen when we normalized on the basis of the timing of P1 cell divisions (Supplementary Fig. 7). These data suggest that local tissue mechanics influence the tendency of cells to delaminate.

To gain insight into the potential role of mechanics in driving cell delamination within a crowded epithelium, we developed a computational model of the notum (details in Supplementary Information) in which the topological organization of cells changed over time as the result of junctional rearrangements while overall tissue area remained fixed (Figs 3a and 4a). First, we used the model to test the likely effects of tissue growth on dynamic cell organization, through the introduction of a variable crowding parameter γ , equal to the ratio of the area that cells would naturally prefer to occupy relative to the fixed tissue area (details in Supplementary Information). In simulations, delamination was directly correlated with crowding, such that high γ favoured delamination by forcing cells to compete for limited space (Fig. 3b). We then examined the effects of cell geometry on delamination by running simulations in which cells either had an isotropic shape or an

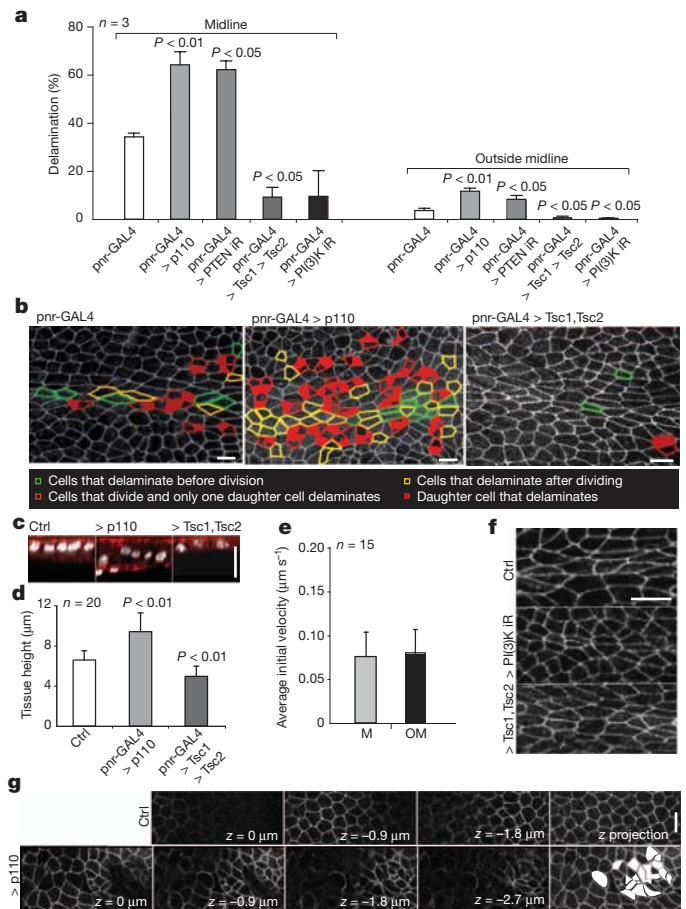


Figure 2 | Delamination counter-balances tissue growth. **a**, Average rates of delamination ($n = 3$ pupae) for cells in the midline and outside midline in control animals, in tissues where growth was increased (p110 PI(3)K, PTEN RNAi) and in nota with decreased growth (p110 PI(3)K RNAi, Tsc1 and Tsc2). **b**, Cells delaminating were coloured according to their behaviour (see inset). **c**, *xz* axis of tissues stained with 4',6-diamidino-2-phenylindole and labelled phalloidin. **d**, Average tissue height ($n = 20$ different regions from 4 different pupae). **e**, Quantification of average initial velocity of vertex displacement after laser dissection of single junctions in early pupae for tissues expressing Tsc1 and Tsc2, with n representing 15 cuts from 5 pupae. **f**, Midline geometry for nota at 14 h AP. **g**, *xy* image planes for different *z*-slices. Delamination is most prevalent in tissue folds after p110 PI(3)K overexpression (coloured white in the *z* projection). All error bars indicate s.d. *P* values were calculated relative to the control. Scale bars, 10 μm .

average initial aspect ratio of 2.7:1, to model cells of the notum midline. This showed that cellular anisotropy promoted delamination (Fig. 3c), which in turn homogenized cell packing (Figs 1a and 3c). Significantly, by combining the two factors, crowding and geometry, we were able to accurately replicate the behaviours of cells within the midline (with anisotropic geometry and increased crowding) and outside the midline (isotropic geometry and no growth-induced crowding) using a single set of fixed parameters (Fig. 3d).

A closer examination of simulations revealed the typical path of cell delamination. Spontaneous fluctuations in the length of individual cell–cell junctions cause a subset of cells to suffer a stochastic serial loss of junctions through neighbour exchange events^{10,11} (Fig. 3e). The concomitant progressive loss in apical area results in the removal of these cells from the simulated tissue. In view of this computational analysis, when we examined the precise path of cell delamination in the developing fly notum we identified two mechanistically distinct paths of delamination. First, we identified a population of cells that leave the tissue over an extended and variable period of time as they undergo a consecutive series of neighbour exchange events^{10,11}, accompanied by a

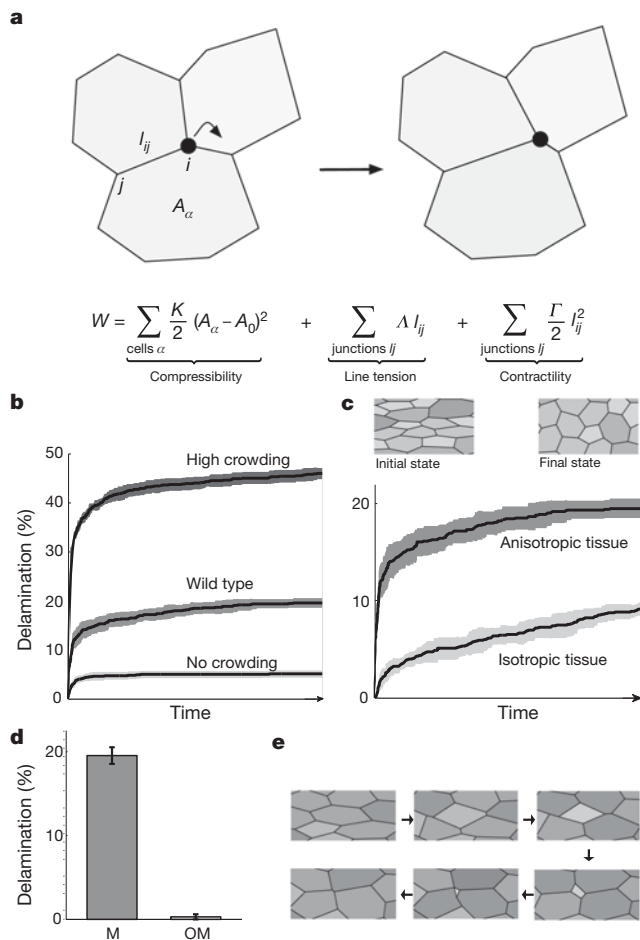


Figure 3 | Topological rearrangements drive a stochastic process of delamination in overcrowded tissues. **a**, Schematic diagram of the model of mechanical forces that act in the plane of the epithelium: junctional line tension (Λ), contractility of cortical acto-myosin (Γ) and limited cell compressibility (K) based on a target area (A_0). The tissue evolves by small stochastic displacements of the vertices, biased towards changes that decrease the work function, W . **b**, Cumulative proportion of cells delaminating from wild-type tissue ($\gamma = 2$) and a tissue with high ($\gamma = 4$) or low ($\gamma = 1$) crowding as a function of time, all three with anisotropic geometry. **c**, Still images showing initial and final cell configurations for the midline simulation (anisotropic); values of parameters correspond to moderate crowding ($\gamma = 2$ in both cases). **d**, Final proportions of cells delaminating in simulations of wild-type tissue within and outside the midline. The midline has anisotropic geometry and $\gamma = 2$; tissue outside the midline has isotropic geometry and $\gamma = 1$. **e**, An example showing the serial loss of junctions and the progressive loss of apical area that precedes cell extrusion. In **b**, **c** and **d** mean values and standard errors of the data are averaged over six simulations. Grey shades represent standard errors of the data.

progressive loss of junctions and apical area (Fig. 4a, b). We term this type of delamination a D1 process. Once these cells had fewer than five neighbours and less than $\sim 25\%$ of their initial area, they were forced out of the tissue within a short and reproducible timescale (about 10 min). This final push was associated with the recruitment of a contractile myosin II ring that formed within neighbouring cells (Fig. 4g). When the actomyosin cable was cut, both cell delamination and delamination-induced cell death were prevented (Supplementary Fig. 8). In addition, we identified a second pathway, which we term a D2 process, by which cells delaminated from the epithelium over a period of about 50 min, as the result of a progressive loss of apical area, without a concomitant change in neighbour relationships (Fig. 4c). This D2 process resembled apoptotic cell extrusion as described previously⁶ in that it was associated with the formation of rosette-like intermediates and the loss of junctional E-cadherin (Fig. 4g and

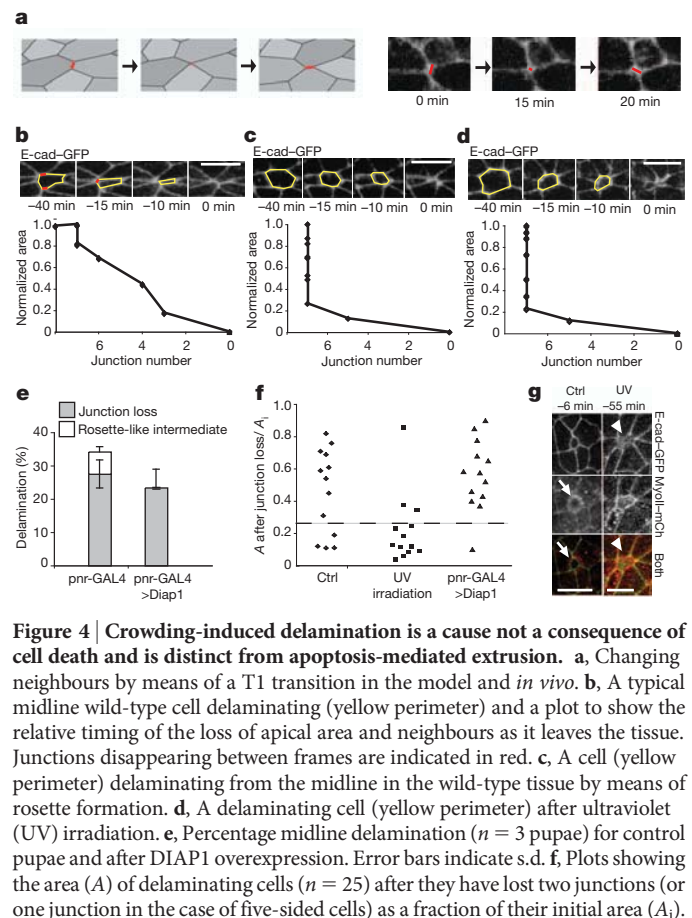


Figure 4 | Crowding-induced delamination is a cause not a consequence of cell death and is distinct from apoptosis-mediated extrusion. **a**, Changing neighbours by means of a T1 transition in the model and *in vivo*. **b**, A typical midline wild-type cell delaminating (yellow perimeter) and a plot to show the relative timing of the loss of apical area and neighbours as it leaves the tissue. Junctions disappearing between frames are indicated in red. **c**, A cell (yellow perimeter) delaminating from the midline in the wild-type tissue by means of rosette formation. **d**, A delaminating cell (yellow perimeter) after ultraviolet (UV) irradiation. **e**, Percentage midline delamination ($n = 3$ pupae) for control pupae and after DIAP1 overexpression. Error bars indicate s.d. **f**, Plots showing the area (A) of delaminating cells ($n = 25$) after they have lost two junctions (or one junction in the case of five-sided cells) as a fraction of their initial area (A_i). Cells above the dotted line delaminate by losing junctions; cells below the dotted line delaminate through the formation of rosette-like intermediates. **g**, Control and ultraviolet-irradiated tissues marked with E-cadherin-GFP and MyoII-Cherry (mCh). The cytoplasmic myosin II ring is indicated with arrows. In the ultraviolet-irradiated tissue the arrowheads indicate E-cadherin loss and junctional myosin II accumulation. Numbers refer to the time in minutes before delamination. Scale bars, 10 μm .

Supplementary Fig. 9). Moreover, this D2 process was actively induced by ultraviolet irradiation (Fig. 4d, f, g). To test whether D1 and/or D2 processes were triggered by dying cells in the wild-type notum, we then monitored cell behaviour in nota that overexpressed an inhibitor of apoptosis, DIAP1 (ref. 12). This led to a small but reproducible decrease in the overall rate of delamination as the result of a selective inhibition of D2-type delamination events (Fig. 4e, f and Supplementary Movie 6). Therefore, although a subset of cells are extruded from the notum as a direct consequence of cell death¹³ by means of a D2-type process, the vast majority of cells leave crowded regions of the tissue while alive, by means of a series of stochastic neighbour exchange events (D1). These live delaminated cells then die in a process akin to anoikis, their nuclei becoming visibly condensed and pyknotic (Supplementary Fig. 10a) before they are cleared from underneath the epithelium by circulating macrophages (Supplementary Fig. 10b). We confirmed that death is a consequence not the cause of D1-type delamination events¹⁴ by showing that midline cell delamination continued unabated in nota expressing RNAi targeting JNK, p53 or Flower (Supplementary Fig. 11) and after p35 expression (data not shown). Finally, we verified that only a subset of cells delaminated from the midline during the entire course of normal development, as expected for a stochastic process (Supplementary Fig. 12).

Taken together, these data identify a previously unknown two-step process of live-cell delamination, which contributes to the refinement of epithelial cell packing at the end of development. We show that this process of crowding-induced delamination can be recapitulated by

using a simple mechanical model of an epithelium in which cells compete for space that includes, first, stochastic junctional fluctuations that are capable of triggering neighbour exchange events, and, second, a final push that can remove cells with small apices from the tissue. Because many epithelia are both crowded and topologically active, we expect crowding-induced delamination to be a common process that buffers tissue growth to aid normal tissue homeostasis in a wide variety of systems. This study also reveals a potential mechanistic link between tissue hyperplasia and delamination, which may have important consequences for our understanding of cancer because it suggests that the spread of a primary epithelial tumour will be limited not by a switch from an epithelial to a mesenchymal state but simply by the tendency of delaminating cells to die¹⁴.

METHODS SUMMARY

For this study, fluorescently marked *pnr-Gal4* pupae expressing different transgenes or RNAi constructs were imaged between 14 h and 26 h AP. Three or more animals were used for each experiment and, when data were quantified, significance was established with an unpaired *t*-test. For ultraviolet irradiation, pupae were dissected at 15 h AP and the exposed notum was irradiated with a Stratalinker 1800 crosslinker at 100 mJ cm⁻², before live imaging. For laser ablation of single junctions at 14 h and 24 h AP, the tissue was imaged with 488-nm light from an Ar-Kr laser every 2 s and ablated with 730-nm multiphoton excitation from a Chameleon-XR Ti-Sapphire laser (AIM, Zeiss).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March 2011; accepted 23 February 2012.

Published online 15 April 2012.

- Hufnagel, L., Teleman, A. A., Rouault, H., Cohen, S. M. & Shraiman, B. I. On the mechanism of wing size determination in fly development. *Proc. Natl Acad. Sci. USA* **104**, 3835–3840 (2007).
- Kafer, J., Hayashi, T., Maree, A. F., Carthew, R. W. & Graner, F. Cell adhesion and cortex contractility determine cell patterning in the *Drosophila* retina. *Proc. Natl Acad. Sci. USA* **104**, 18549–18554 (2007).
- Farhadifar, R., Roper, J. C., Aigouy, B., Eaton, S. & Julicher, F. The influence of cell mechanics, cell–cell interactions, and proliferation on epithelial packing. *Curr. Biol.* **17**, 2095–2104 (2007).
- Aegerter-Wilmsen, T. *et al.* Exploring the effects of mechanical feedback on epithelial topology. *Development* **137**, 499–506 (2010).
- Cohen, M., Georgiou, M., Stevenson, N. L., Miodownik, M. & Baum, B. Dynamic filopodia transmit intermittent Delta–Notch signaling to drive pattern refinement during lateral inhibition. *Dev. Cell* **19**, 78–89 (2010).
- Rosenblatt, J., Raff, M. C. & Cramer, L. P. An epithelial cell destined for apoptosis signals its neighbors to extrude it by an actin- and myosin-dependent mechanism. *Curr. Biol.* **11**, 1847–1857 (2001).
- Toyama, Y., Peralta, X. G., Wells, A. R., Kiehart, D. P. & Edwards, G. S. Apoptotic force and tissue dynamics during *Drosophila* embryogenesis. *Science* **321**, 1683–1686 (2008).
- Manjon, C., Sanchez-Herrero, E. & Suzanne, M. Sharp boundaries of Dpp signalling trigger local cell death required for *Drosophila* leg morphogenesis. *Nature Cell Biol.* **9**, 57–63 (2007).
- Zeitlinger, J. & Bohmann, D. Thorax closure in *Drosophila*: involvement of Fos and the JNK pathway. *Development* **126**, 3947–3956 (1999).
- Fristrom, D. The cellular basis of epithelial morphogenesis. A review. *Tissue Cell* **20**, 645–690 (1988).
- Irvine, K. D. & Wieschaus, E. Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development* **120**, 827–841 (1994).
- Brand, A. H., Manoukian, A. S. & Perrimon, N. Ectopic expression in *Drosophila*. *Methods Cell Biol.* **44**, 635–654 (1994).
- Koto, A., Kuranaga, E. & Miura, M. Apoptosis ensures spacing pattern formation of *Drosophila* sensory organs. *Curr. Biol.* **21**, 278–287 (2011).
- Emoto, Y. Cellular aggregation facilitates anoikis in MDCK cells. *J. Physiol. Sci.* **58**, 371–380 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Rosenblatt for communicating results before publication; N. Tapon, F. Pichaud, G. Charras and J. Rohn for helpful comments on the text; and the members of the Baum laboratory, in particular K. Van Hegan, J. Bellis and J. Beira. B.B. and E.M. thank Cancer Research UK for funding. B.B. also thanks University College London, Wellcome and the Royal Society for financial support.

Author Contributions E.M. conducted the experiments detailed in the paper, aided by S.C. The laser ablation work was done with technical assistance from J.G. A.M. performed the theoretical analysis. T.D. oversaw the theoretical analysis. B.B. oversaw the experimental work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.B. (b.baum@ucl.ac.uk).

METHODS

Fly stocks and genetics. The following stocks were used: *pnr-GAL4* (Bloomington:3039), *E-cadherin-GFP* (*ubi-DE-cad-GFP*)¹⁵, *E-cadherin-GFP His-RFP* (III) (*ubi-DE-cad-GFP/CyO;His-RFP/TM6B*), *Act-5c-GAL4* (Bloomington:25374), p110 overexpression (*w;UAS-myc-Dp110*)¹⁶, DIAP1 overexpression (III) (*ijf/CyO;UAS-DIAP1*), Tsc1 and Tsc2 overexpression (III) (*hsFLP;UAS-Tsc1,Tsc2*), *Resille::GFP*¹⁷. *E-cadherin-GFP* was crossed with the *pnr-GAL4* line (*ubi-DE-cad-GFP;pnr-GAL4/TM6B*) to drive gene or IR expression. For IR expression, males from each *UAS-IR* fly line were crossed with *Ecad-GFP; pnr-GAL4* virgin females. Crosses between *UAS-hairpin RNAi* males and *Ecadherin-GFP; pnr-GAL4* females were raised at 22 °C and shifted to 29 °C during larval stages. RNAi lines were used to silence the expression of the following genes: *cdc25* (ID17760), *fwe* (ID104993), *bsk* (ID34138), *pl3k* (ID38985), *pten* (ID35731), *r5 s* (ID27792), *warts* (ID9928). All RNAi lines came from the Vienna *Drosophila* RNAi Center (VDRC) library¹⁸.

Live imaging and dissections. Live imaging¹⁹ was performed by cutting a window in the pupal case attached to a slide with double-sided tape, and placing a coverslip carrying a drop of injection oil over the notum, supported by coverslips at either end. Notum were imaged from 14 h AP for 12 h during development. For stained tissues, notum from 16 h AP pupae were dissected in PBS for direct fixation in 4% formaldehyde for 20 min before being permeabilized with PBS containing 0.1% Triton X-100 (PBT). The notum were then stained for 2 h at 25 °C with 4',6-diamidino-2-phenylindole (Sigma) for nuclei and Alexa-568 phalloidin (Invitrogen) for actin. The acquisition was performed with a Leica SP2 laser scanning confocal microscope with 40×/1.3 numerical aperture oil objective for both fixed and live-imaging experiments; three or more animals were used for each experiment and, when data were quantified, significance was established with an unpaired *t*-test in Excel.

Ultraviolet irradiation. Pupae were dissected at 15 h AP and the exposed notum was irradiated with a Stratalinker 1800 crosslinker at 100 mJ cm⁻², before live imaging.

Laser ablation. Single junctions were ablated in pupae at 14–16 h (early pupae) and 24–26 h (late pupae), using 730-nm multiphoton excitation from a Chameleon-XR Ti:Sapphire laser, under software control (AIM, Zeiss). A 3 × 3-pixel (pixel size 0.17 μm) region of the junction was irradiated at slow scan speed (pixel dwell time 2.56 μs) at 25% laser power, under software control. The power required to ablate junctions was determined empirically each time and was used in each experiment (AIM, Zeiss). Ablations were triggered during time-lapse imaging (every 2 s), using 488-nm light from an Ar–Kr laser coupled to a Zeiss 510NLO upright confocal microscope and a 40×/1.3 numerical aperture oil objective.

Image processing and analysis. The images presented were processed with ImageJ (<http://rsb.info.nih.gov/ij/>) and Adobe Photoshop CS (Adobe Systems, Inc.). Unless otherwise noted, images of the tissue represent maximum intensity *z*-projections, and sections were 0.9 μm apart. For laser ablation, the displacement of vertices was measured after a cut with the ImageJ Kymograph plugin. Statistics were calculated in each case on the basis of 15 junctions in 5 different pupae.

15. Oda, H. & Tsukita, S. Real-time imaging of cell–cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114**, 493–501 (2001).
16. Leever, S. J., Weinkove, D., MacDougall, L. K., Hafen, E. & Waterfield, M. D. The *Drosophila* phosphoinositide 3-kinase Dp110 promotes cell growth. *EMBO J.* **15**, 6584–6594 (1996).
17. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
18. Dietzl, G. et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
19. Georgiou, M., Marinari, E., Burden, J. & Baum, B. Cdc42, Par6, and aPKC regulate Arp2/3-mediated endocytosis to control local adherens junction stability. *Curr. Biol.* **18**, 1631–1638 (2008).

Restoration of grasp following paralysis through brain-controlled stimulation of muscles

C. Ethier¹, E. R. Oby¹, M. J. Bauman² & L. E. Miller^{1,3,4}

Patients with spinal cord injury lack the connections between brain and spinal cord circuits that are essential for voluntary movement. Clinical systems that achieve muscle contraction through functional electrical stimulation (FES) have proven to be effective in allowing patients with tetraplegia to regain control of hand movements and to achieve a greater measure of independence in daily activities^{1,2}. In existing clinical systems, the patient uses residual proximal limb movements to trigger pre-programmed stimulation that causes the paralysed muscles to contract, allowing use of one or two basic grasps. Instead, we have developed an FES system in primates that is controlled by recordings made from micro-electrodes permanently implanted in the brain. We simulated some of the effects of the paralysis caused by C5 or C6 spinal cord injury³ by injecting rhesus monkeys with a local anaesthetic to block the median and ulnar nerves at the elbow. Then, using recordings from approximately 100 neurons in the motor cortex, we predicted the intended activity of several of the paralysed muscles, and used these predictions to control the intensity of stimulation of the same muscles. This process essentially bypassed the spinal cord, restoring to the monkeys voluntary control of their paralysed muscles. This achievement is a major advance towards similar restoration of hand function in human patients through brain-controlled FES. We anticipate that in human patients, this neuroprosthesis would allow much more flexible and dexterous use of the hand than is possible with existing FES systems.

Worldwide, over 130,000 people each year survive spinal cord injury (SCI) but sustain extensive paralysis⁴. Approximately half of these injuries occur above the sixth cervical vertebra, thereby affecting all four limbs. Most of these patients indicate that regaining the ability to grasp objects would provide the greatest practical benefit compared to regaining other lost functions⁵.

For this reason, considerable effort has been devoted to the development of FES systems to restore voluntary grasp^{1,2,6}. These systems rely on residual movement or muscle activity to control electrical activation of hand muscles. Because of the complexity of the necessary patterns of muscle activation, current FES systems produce only one or two grasps using pre-programmed stimulus trains that must be customized for each user⁷. This is effective because many objects can be grasped adequately with only palmar or pinch grasp. However, normal hand use is much more complex than this. Furthermore, using the motion of one body part to control that of another inevitably increases the associated cognitive burden. If FES is to provide hand movements that are close to normal, a more natural control signal of higher dimension than that available through residual motion will be necessary.

Fortunately, the rapid development of the brain machine interface (BMI) provides promising new means by which more flexible and dexterous movements might be controlled. However, despite the initial demonstration of strong force-related discharge in the primary motor

cortex (M1)⁸, virtually all existing BMIs extract only kinematic information from the brain. This bias is ironic, as the first study to decode signals from simultaneously recorded neurons found that force was more strongly represented than movement in M1 (ref. 9).

Only a small number of groups have pursued the possibility of using kinetic (force-related) information as a real-time control signal for a BMI, through the prediction of grasp force^{10,11}, joint torque¹² or muscle activity^{10,13,14}. We showed previously that despite paralysis produced by peripheral nerve block, monkeys could accurately modulate the magnitude of isometric flexion and extension wrist torque using cortically controlled FES^{15,16,17}. Related results were also reported by a group that operantly conditioned monkeys to modulate the activity of one or two individual neurons whose discharge directly controlled stimulation of individual muscles¹⁸.

We performed the current experiments with two monkeys trained to pick up weighted rubber balls and to convey them to an opening at the top of a dispenser (Fig. 1). After training, each monkey was implanted with a multi-electrode recording array in the hand area of M1. In a separate surgical procedure, we implanted intramuscular electrodes for recording and stimulation of hand and forearm muscles. Figure 2a shows the neural discharge recorded under normal conditions from a representative session. Most of these 104 neuronal signals

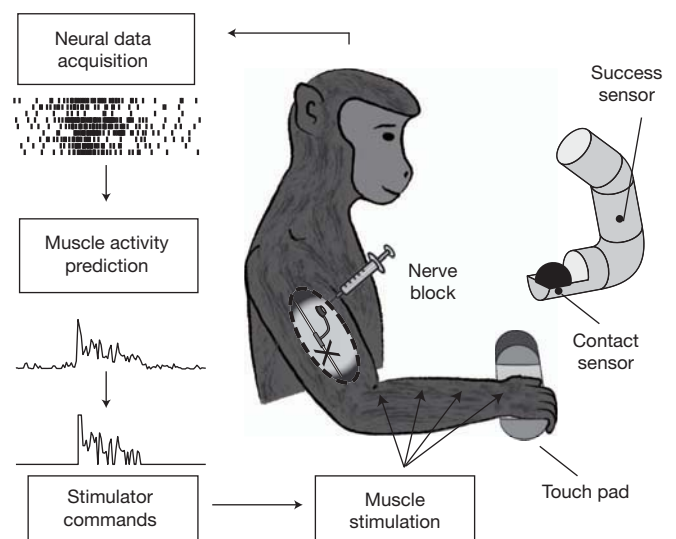


Figure 1 | Brain-controlled FES. The monkeys' forearm and digit flexor muscles were temporarily paralysed by a peripheral nerve block. Recordings from the motor cortex were used to infer the monkeys' attempted patterns of muscle activity and thereby control electrical stimulation that restored the monkeys' ability to perform a functional grasping task. The ball-grasp device was equipped with a contact sensor and a task-success sensor that were activated when the monkeys initially touched the ball and dropped it into the tube, respectively.

¹Department of Physiology, Feinberg School of Medicine, Northwestern University, 303 East Chicago Avenue, Chicago, Illinois 60611, USA. ²Department of Bioengineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. ³Department of Biomedical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA. ⁴Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, 345 East Superior Avenue, Chicago, Illinois 60611, USA.

were well-modulated during at least some portion of the task. Offline, typically 50–75% of the neuronal signals could be discriminated as single neurons, on the basis of the consistency of their waveform shape and inter-spike interval histogram distribution. However, under real-time conditions, only about one-third of the inputs were well-discriminated single units; the remainder were signals that included action potentials from more than one neuron. Figure 2b shows the discharge of these neurons averaged over 229 trials, aligned to the time of contact with the ball. The varied phasing of the different neurons is evident.

We recorded from flexor and extensor muscles of the hand and forearm simultaneously with the neural recordings (Fig. 2c, d). There was considerable variation both in the magnitude and duration of electrical activity that occurred from trial to trial (Fig. 2c), and in the average timing and patterns of activation of the different muscles (Fig. 2d).

We were able to predict electromyographic (EMG) activity with very high accuracy, typically from approximately 100 neural signals (Fig. 2c, d; red traces), using Wiener cascade decoders. These decoders consisted of multiple-input, linear-impulse response functions between the neural inputs and each muscle, followed by a static non-linearity. Each impulse response was composed of ten lags spanning 500 ms. At the beginning of each week, we collected 20 min of data under normal conditions, and we used this to compute the coefficients for the decoder that were then used for the remainder of the week. Accuracy was represented by R^2 , calculated using a multi-fold cross-validation procedure described in the Supplementary Information.

Using these real-time predictions of muscle activity, we stimulated up to five electrodes in three different muscles (flexor carpi radialis

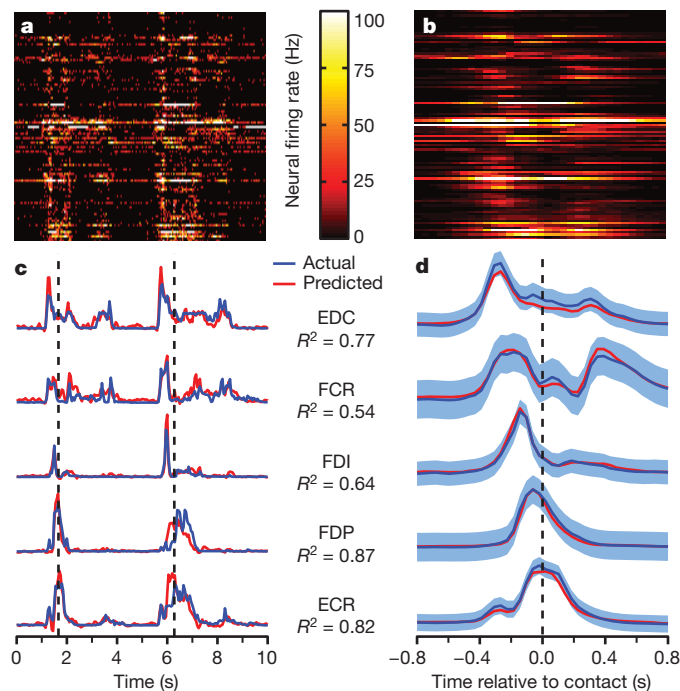


Figure 2 | Grasp-related raw data collected during normal conditions. **a**, Firing rates of 104 neuronal signals recorded during series of two grasps. **b**, Ensemble average of 229 trials aligned to the time of ball contact. **c**, Actual and predicted EMG during the same period as (**a**), with the muscles ordered by the relative times of their onset, including extensor digitorum communis (EDC), flexor carpi radialis (FCR), first dorsal interosseous (FDI), flexor digitorum profundus (FDP), and extensor carpi radialis (ECR). Predicted EMG was computed using multiple-input linear-impulse response decoders built from data collected earlier in the session. Vertical dashed lines mark the times of ball contact. R^2 values indicate prediction accuracy for the 20-min data file. **d**, Ensemble averages of EMG activity, aligned to the time of initial contact. Blue shaded regions, ± 1 s.d. around the mean.

(FCR), medial and lateral sites in the flexor digitorum superficialis (FDS) and flexor digitorum profundus (FDP)). By these means, we have restored in two monkeys the ability to pick up and move objects despite complete paralysis of the flexor muscles in the forearm and hand. We began each FES experimental session by collecting data under normal conditions to establish baseline performance. Following these baseline recordings, we injected lidocaine through nerve cuffs implanted proximal to the elbow that blocked the median and ulnar nerves. After 15–20 min the nerve block was complete, as determined by the loss of flexor muscle EMG activity (see Supplementary Information and Supplementary Fig. 1), and the onset of profound motor deficits. We made periodic tests of nerve-block effectiveness throughout each session (Supplementary Fig. 2), and we used a standardized stimulus train to evaluate the level of fatigue induced by the stimulation (Supplementary Fig. 2).

A series of four trials is shown in Fig. 3a, b, showing typical neural discharge, predicted EMG and stimulus commands, as well as markers of the monkey's performance. Although the common digit flexors (FDS and FDP) are normally activated nearly synchronously, FDS activation tended to be more sustained, whereas FDP was more phasic. The pulse widths of the stimulus trains used to activate a given muscle

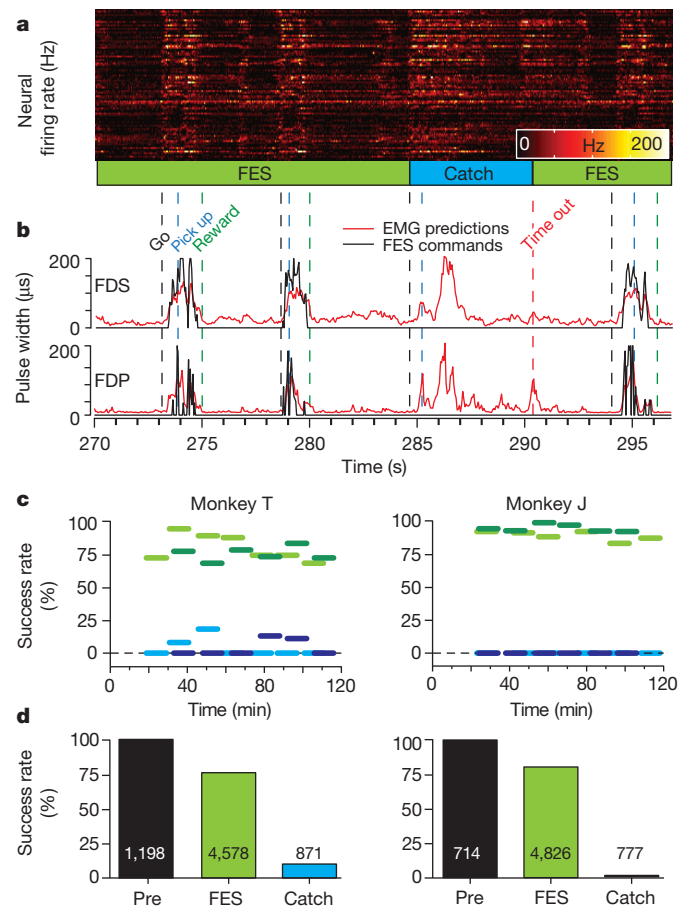


Figure 3 | Grasp performance during four consecutive brain-controlled FES trials. **a**, Neural data. **b**, Predicted EMG signals (red traces) transformed into stimulus commands (black traces). Vertical dashed lines: go tone (Go), time of initial ball contact (Pick up) and successful task completion (Reward). **c**, Horizontal lines show average success rates for sequential 10-min blocks during two separate experimental sessions (indicated by light and dark horizontal lines, respectively). Each session included both FES trials (green lines) and catch trials without stimulation (blue lines). The neuroprosthesis markedly improved the monkeys' ability to grasp the ball despite paralysis. **d**, Average success rates for pre-block (Pre), FES and catch (Catch) trials across all sessions (100%, 76% and 10% for Monkey T; 99%, 80% and 1% for Monkey J). The total number of trials (successful and failed) is displayed on the bars for each condition.

were determined from the predicted EMG for that muscle using a mapping procedure described in the Supplementary Information and Supplementary Fig. 3. The distribution of these pulse widths throughout the full range from 0 to 200 μ s suggests that the monkey was able to grade the strength of contraction continuously (Supplementary Fig. 4). During the FES trials, the monkey grasped and moved the ball reliably. The movements did not differ sufficiently from normal to be obvious to casual observation (see Supplementary Movies 1 and 2 for representative examples from both monkeys). On occasional 'catch' trials, we turned off the neuroprosthesis at the beginning of the trial, to test the ability of the monkey without FES. In the example of a catch trial illustrated here (note the flat stimulus trace in Fig. 3b), the monkey was unable to grasp the ball despite the considerable effort apparent in the neural discharge and predicted EMG.

After the onset of paralysis, each experimental session consisted of a series of 10-min sets of trials like those in Fig. 3, in which the monkey attempted to complete the grasp task either with or without FES assistance. Two complete sessions for both monkeys are summarized by the horizontal light and dark lines in Fig. 3c. The success rate in these sessions using the neuroprosthesis was approximately 80% and 90% for the two monkeys, respectively (green lines). In contrast, the average catch-trial success rate was 5% for monkey T and 0% for monkey J (blue lines). The average number of trials per session varied substantially across sessions, with a mean of 272 ± 84 for monkey T, and 208 ± 112 for monkey J. Although we tried different types of balls, we did not systematically examine the effects of size, weight or texture on the monkeys' performance. It is likely that the FES success rate would have been lower if balls that were substantially heavier or more slippery had been used. We chose to use balls that in size and weight mimicked objects grasped in routine human tasks (for example, eating an apple).

Figure 3d summarizes both monkeys' overall success rate across all sessions, both with the FES neuroprosthesis and during catch trials. Both monkeys achieved a success rate of approximately 80% using the neuroprosthesis, a level that was highly significantly different ($P < 0.0001$) from that of the catch-trial condition. In addition to resulting in a greatly improved success rate, the FES neuroprosthesis also significantly increased the speed at which the monkeys completed successful trials (not shown; $P < 0.0001$ for both monkeys, two-tailed Mann-Whitney test).

To test force control more systematically we conducted a second set of experiments with monkey J, who was trained to control the vertical displacement of a cursor that moved in proportion to palmar grasp force. Using the neuroprosthesis, the monkey was able to squeeze a pneumatic tube, and to track up to three different targets ranging from 15 to 80% of his normal maximum voluntary contraction (MVC), each target having a width of approximately 20% of MVC. To be successful, the monkey needed to maintain the target force for 0.5 s. Figure 4 shows a short sequence of data during this target tracking task. One of these four trials was a catch trial. The monkey was unable to generate any force during the catch trial despite two attempts that are evident in the predicted EMG signals.

We quantified this performance by measuring the mean force and stimulation pulse width during the target-hold periods of the initial and final 10 min of the session. Despite considerable FES-induced fatigue, the monkey remained able to achieve the required force throughout the session by voluntarily increasing the mean stimulus pulse width (see Supplementary Fig. 5). The increased pulse width reflects an increase in cortical activity and resultant EMG predictions. The monkey seemed to overcome the fatigue in a manner similar to that of normal conditions, increasing its effort to regulate force accurately.

The monkey's ability to control both a well-regulated palmar grasp as well as to execute the unconstrained natural grasp is powerful evidence of the impact that this FES neuroprosthesis could have in eventual clinical application. Our neuroprosthesis makes use of patterns of activity in M1 that reflect the patterns that occur naturally during grasp. By matching patterns of neuronal activity to those muscles with

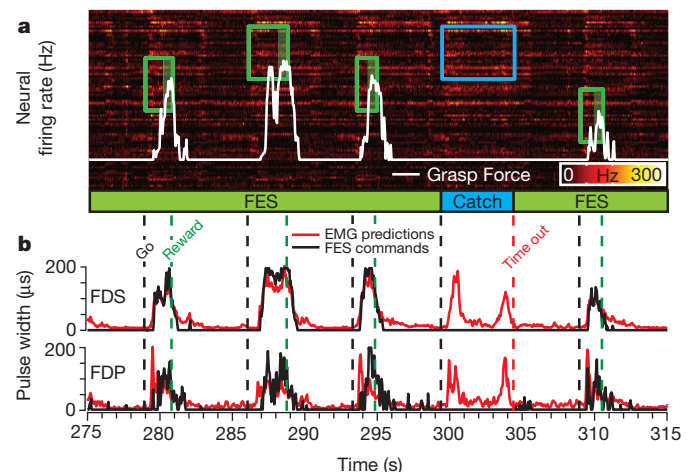


Figure 4 | FES used to produce controlled palmar grasp force during the palmar grasp task. **a**, Colour-coded neural discharge recorded during a series of five trials. The rectangles indicate times of target appearance and disappearance, as well as their upper and lower force bounds. The white trace is the force generated by the monkey, resulting from stimulation of FDS and FDP (Stim commands). **b**, There were four successful trials with FES (green targets) and one unsuccessful catch trial (blue target). During the catch trial, the monkey made two unsuccessful attempts to squeeze the tube, as seen in the EMG predictions (also evident in the neural discharge (**a**)).

which they are normally most closely correlated, we hope to maintain the natural coupling between cortical activity and motor output.

It is important to note that this process in no way limits the ability of the brain to adapt further, to compensate for inaccuracies in the decoded signals or the stimulus-evoked contractions. Even with adaptation, it is difficult to imagine how a small number of individually conditioned, randomly selected neurons could yield an adequate level of control without the type of pre-programming that is necessary with existing FES systems. Indeed, there is no evidence that it is possible to learn to associate the simultaneous activity of two, three or more neurons with independent patterns of muscle activity. Even if possible, the cognitive load associated with this effort would presumably be rather high, whereas the reliability of a neuroprosthesis relying on a small number of conditioned neurons would be quite low.

Our model of paralysis avoided many of the complications of actual spinal cord injury, including muscle denervation and spasticity^{19,20}. Furthermore, it was limited to the forearm and digit flexors. Patients with C5 and C6 spinal cord injury retain voluntary control of proximal arm muscles while losing full control of the more distal limb. Many patients retain or regain some level of voluntary wrist extension²¹. As we did not paralyse the monkeys' extensor muscles in this experiment, it is important to recognize the good coordination between the remaining natural muscle control and that achieved through the neuroprosthesis. We routinely obtained extensor EMG predictions that were in fact slightly more accurate than those of the flexors. In future experiments, we intend to expand our control to these muscles.

This technology may offer even greater advantages to patients with more severe injuries, who have a greater need for replaced function but possess even fewer available sources of control²². In addition to the ability to predict the reach-related activity of distal limb muscles considered in this study, we have previously showed the same ability in relation to proximal limb muscles, suggesting the possibility of extending this control to these muscles¹³. As well as offering patients greater independence, FES is also established as an effective means for exercising the muscles of paralysed patients, bringing a range of health benefits: stronger muscles and bones, improved metabolism, cardiorespiratory health and reduced propensity to pressure sores^{23,24}. It may be that drawing on a conscious process to restore natural movement will bring the additional benefit of improved psychological health²⁵.

METHODS SUMMARY

Experimental subjects and task. Two monkeys were trained to perform a ball-grasp task (Fig. 1) and one of the monkeys was also trained to perform a controlled-force palmar grasp task. The monkeys were allowed 5 s to grasp one of several balls (ranging in size from 25–40 mm in diameter and 55–130 g) and place it into the top of a dispenser tube. The palmar grasp task required the monkey to squeeze a pneumatic tube that controlled movement of a cursor. Force targets were chosen from a set of two or three non-overlapping levels. All procedures were approved by the Institutional Animal Care and Use Committee of Northwestern University, Illinois, USA.

EMG prediction. Inputs consisted of roughly 100 single and multi-unit signals from a 100-electrode array (Blackrock Microsystems) implanted within the hand area of M1. Decoders consisted of multiple-input impulse response functions between the neural inputs and each muscle, subsequently transformed by a second-order static nonlinearity to reduce the baseline noise in the predictions and to increase the gain near the EMG peaks^{13,16}. We computed decoders at the beginning of each week, which were used in daily sessions for the remainder of the week. We conducted 20 sessions with 7 decoders across 7 weeks for monkey T and 27 sessions with 6 decoders across 11 weeks with monkey J.

Stimulation. All muscles were stimulated at a single, fixed rate of either 25 or 30 Hz to achieve nearly fused contractions. The EMG predictions were transformed into stimulus pulse widths by mapping the predicted EMG noise floor to the stimulus force threshold, and the maximum predicted EMG to the maximum pulse width (200 μ s; see Supplementary Fig. 3). The current, typically 2–8 mA, was chosen independently for each electrode, to yield forces of approximately 50% of the maximal evocable force at 200- μ s pulse width.

Received 23 December 2011; accepted 22 February 2012.

Published online 18 April 2012.

- Keith, M. W. *et al.* Implantable functional neuromuscular stimulation in the tetraplegic hand. *J. Hand Surg. Am.* **14**, 524–530 (1989).
- Peckham, P. H. *et al.* Efficacy of an implanted neuroprosthesis for restoring hand grasp in tetraplegia: a multicenter study. *Arch. Phys. Med. Rehabil.* **82**, 1380–1388 (2001).
- Pohlmeier, E. A., Jordon, L. R., Kim, P. & Miller, L. E. A fully implanted drug delivery system for peripheral nerve blocks in behaving animals. *J. Neurosci. Methods* **182**, 165–172 (2009).
- International. Campaign for Cures of Spinal Cord Injury Paralysis. <http://www.campaignforcure.org> (2011).
- Anderson, K. D. Targeting recovery: priorities of the spinal cord-injured population. *J. Neurotrauma* **21**, 1371–1383 (2004).
- Popovic, M. R., Popovic, D. B. & Keller, T. Neuroprostheses for grasping. *Neurol. Res.* **24**, 443–452 (2002).
- Kilgore, K. L. *et al.* An implanted upper-extremity neuroprosthesis. Follow-up of five patients. *J. Bone Joint Surg. Am.* **79**, 533–541 (1997).
- Evarts, E. V. Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* **31**, 14–27 (1968).
- Humphrey, D. R., Schmidt, E. M. & Thompson, W. D. Predicting measures of motor performance from multiple cortical spike trains. *Science* **170**, 758–761 (1970).
- Carmena, J. M. *et al.* Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, e42 (2003).
- Gupta, R. & Ashe, J. Offline decoding of end-point forces using neural ensembles: application to a brain-machine interface. *Neural Systems and Rehabilitation Engineering. IEEE Trans. Neural Syst. Rehabil. Eng.* **17**, 254–262 (2009).
- Fagg, A. H., Ojakangas, G. W., Miller, L. E. & Hatsopoulos, N. G. Kinetic trajectory decoding using motor cortical ensembles. *IEEE Trans. Neural Syst. Rehabil. Eng.* **17**, 487–496 (2009).
- Pohlmeier, E. A., Solla, S. A., Perreault, E. J. & Miller, L. E. Prediction of upper limb muscle activity from motor cortical discharge during reaching. *J. Neural Eng.* **4**, 369–379 (2007).
- Cherian, A., Krucoff, M. O. & Miller, L. E. Motor cortical prediction of EMG: evidence that a kinetic brain-machine interface may be robust across altered movement dynamics. *J. Neurophysiol.* **106**, 564–575 (2011).
- Pohlmeier, E. A. *et al.* Real-time control of the hand by intracortically controlled functional neuromuscular stimulation. *IEEE 10th Int. Conf. Rehab. Robotics* 454–458 (2007).
- Pohlmeier, E. A. *et al.* Toward the restoration of hand use to a paralyzed monkey: Brain-controlled functional electrical stimulation of forearm muscles. *PLoS ONE* **4**, e5924 (2009).
- Oby, E. R. *et al.* in *Statistical Signal Processing for Neuroscience and Neurotechnology* (ed. O'Weiss, K.G.) 369–406 (Academic Press, 2010).
- Moritz, C. T., Perlmutter, S. I. & Fetz, E. E. Direct control of paralysed muscles by cortical neurons. *Nature* **456**, 639–642 (2008).
- Adams, M. M. & Hicks, A. L. Spasticity after spinal cord injury. *Spinal Cord* **43**, 577–586 (2005).
- Kern, H. *et al.* Denervated muscles in humans: limitations and problems of currently used functional electrical stimulation training protocols. *Artif. Organs* **26**, 216–218 (2002).
- Waters, R. L., Adkins, R. H., Yakura, J. S. & Sie, I. Motor and sensory recovery following complete tetraplegia. *Arch. Phys. Med. Rehabil.* **74**, 242–247 (1993).
- Bryden, A. M. *et al.* An implanted neuroprosthesis for high tetraplegia. *Top. Spinal Cord Inj. Rehabil.* **10**, 38–52 (2005).
- Nightingale, E. J., Raymond, J., Middleton, J. W., Crosbie, J. & Davis, G. M. Benefits of FES gait in a spinal cord injured population. *Spinal Cord* **45**, 646–657 (2007).
- Agarwal, S. *et al.* Long-term user perceptions of an implanted neuroprosthesis for exercise, standing, and transfers after spinal cord injury. *J. Rehabil. Res. Dev.* **40**, 241–252 (2003).
- Fitzwater, R. A personal user's view of functional electrical stimulation cycling. *Artif. Organs* **26**, 284–286 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by grant NS053603 from the National Institute of Neurological Disorders and Stroke to L.E.M. and a post-doctoral fellowship from the Fonds de la Recherche en Santé du Québec to C.E., with further support from the Chicago Community Trust through the Searle Program for Neurological Restoration at the Rehabilitation Institute of Chicago. We also acknowledge the technical assistance of D. Tyler and K. Kilgore as well as the surgical assistance of J. Ko and S. Paisley Agnew.

Author Contributions L.E.M. conceived, designed and supervised the basic experiments. C.E. and E.R.O. performed the experiments. M.J.B. carried out software development. C.E. analysed the data and prepared figures. L.E.M. and C.E. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.E.M. (lm@northwestern.edu).

Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen

Liya Hu¹, Sue E. Crawford², Rita Czakó², Nicolas W. Cortes-Penfield², David F. Smith³, Jacques Le Pendu^{4,5,6}, Mary K. Estes² & B. V. Venkataram Prasad^{1,2}

As with many other viruses, the initial cell attachment of rotaviruses, which are the major causative agent of infantile gastroenteritis, is mediated by interactions with specific cellular glycans^{1–4}. The distally located VP8* domain of the rotavirus spike protein VP4 (ref. 5) mediates such interactions. The existing paradigm is that ‘sialidase-sensitive’ animal rotavirus strains bind to glycans with terminal sialic acid (Sia), whereas ‘sialidase-insensitive’ human rotavirus strains bind to glycans with internal Sia such as GM1 (ref. 3). Although the involvement of Sia in the animal strains is firmly supported by crystallographic studies^{1,3,6,7}, it is not yet known how VP8* of human rotaviruses interacts with Sia and whether their cell attachment necessarily involves sialoglycans. Here we show that VP8* of a human rotavirus strain specifically recognizes A-type histo-blood group antigen (HBGA) using a glycan array screen comprised of 511 glycans, and that virus infectivity in HT-29 cells is abrogated by anti-A-type antibodies as well as significantly enhanced in Chinese hamster ovary cells genetically modified to express the A-type HBGA, providing a novel paradigm for initial cell attachment of human rotavirus. HBGA are genetically determined glycoconjugates present in mucosal secretions, epithelia and on red blood cells⁸, and are recognized as susceptibility and cell attachment factors for gastric pathogens like *Helicobacter pylori*⁹ and noroviruses¹⁰. Our crystallographic studies show that the A-type HBGA binds to the human rotavirus VP8* at the same location as the Sia in the VP8* of animal rotavirus, and suggest how subtle changes within the same structural framework allow for such receptor switching. These results raise the possibility that host susceptibility to specific human rotavirus strains and pathogenesis are influenced by genetically controlled expression of different HBGA among the world’s population.

Rotaviruses are classified on the basis of the neutralization specificity of the outer capsid proteins VP7 and VP4 into G (VP7) and P (VP4) genotypes following a dual nomenclature system similar to influenza viruses¹¹. The crystallographic structures of VP8* from two sialidase-insensitive human strains, representing P[8] (Wa)¹ and P[4] (DS1)¹², from two sialidase-sensitive animal strains, representing P[3] (RRV)^{6,7} and P[7] (CRW-8)¹, and the structures of two animal VP8* with bound Sia^{1,6,12} have been previously reported. NMR, cell-binding and neutralization assays showed that the sialidase-insensitive P[8] Wa strain binds to gangliosides such as GM1 using internal Sia³. These studies suggested that whereas the sialidase-sensitive strains recognize glycans with terminal Sia such as GD1a, the sialidase-insensitive rotavirus strains bind to gangliosides such as GM1 with an internal Sia moiety, and gave rise to the notion that Sia is the key determinant for host-cell recognition in rotaviruses. Our goal was to determine whether all sialidase-insensitive human rotavirus genotypes recognize gangliosides with an internal Sia moiety for initial cell attachment, or

whether they recognize different glycans in a genotype-dependent manner. VP8* (amino acids 64–224), cloned from a human rotavirus strain (HAL1166) first isolated from a child in Finland¹³, was expressed in *Escherichia coli*, purified to homogeneity and crystallized for structural analysis. The sialidase-insensitive HAL1166 strain, phylogenetically and serologically belongs to G8P[14] genotype¹⁴. Although not as prevalent as the P[4] and P[8] genotypes, the P[14] genotypes are being increasingly documented by global rotavirus surveillance^{15–17}, and P[14] human rotaviruses are thought to be able to jump from animal to human hosts¹⁷.

The structure of the HAL1166 VP8* determined to 1.5 Å resolution shows the characteristic galectin-like fold with two twisted β-sheets separated by a shallow cleft as observed in the VP8* structures from other rotavirus strains (Fig. 1a). The structure of P[14] VP8* superimposes well with all of the VP8* structures previously determined. One significant difference between these structures is in the width of the cleft separating the two twisted β-barrel sheets (Fig. 1b). In P[14] VP8*, it is narrower than the cleft in the VP8* of the other two human strains, similar to that in the VP8* of the animal strains. In the animal VP8* structures, Sia binds near the cleft (Fig. 1c). Although the cleft in the P[14] VP8* structure is of similar dimensions as in the animal VP8* structures, the structural features in this region of P[14] VP8* is not compatible with Sia binding. In addition to changes in the amino acid residues (Fig. 1d) and side-chain orientations, the positioning of the amino acid residues is slightly shifted in this region because of an insertion (amino acid 187). The side-chain of Y188 is oriented in such a way that it would cause steric hindrance to Sia binding (Fig. 1c). Furthermore, the P[14] VP8* structure with a narrower cleft and several amino acid changes (Fig. 1d) is not compatible with binding of GM1 as suggested, based on computer modelling, for VP8* of the other human strains with a wider cleft³.

These observations prompted us to undertake a high throughput screening of a glycan array comprised of 511 different glycans, including several glycans with terminal or internal Sia. Such screening, which has been used to identify cellular glycans for a variety of pathogens including bacterial toxins¹⁸, influenza viruses¹⁹ and polyomavirus²⁰, unambiguously showed specific binding to glycans with a terminal GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAc sequence (Fuc, fucose; Gal, galactose; Glc, glucose; NAc, N-acetyl, which is a characteristic of A-type HBGA (Supplementary Table 1, Supplementary Figs 1 and 2). None of the sialylated glycans, with either internal or terminal Sia, showed significant binding (Supplementary Table 1, blue).

To understand the structural interactions between P[14] VP8* and A-type HBGA, we co-crystallized VP8* with tri- and tetrasaccharides that correspond to the terminal structure in the A-type HBGA. The structure of the complex, determined to a similar resolution of 1.5 Å as the unliganded structure, clearly showed density for the bound ligand

¹Verna and Marrs McLean, Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. ³Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322, USA. ⁴INSERM, UMR892, Université de Nantes, 44007 Nantes, France. ⁵CNRS, UMR 6299, Université de Nantes, 44007 Nantes, France. ⁶Université de Nantes, 44007 Nantes, France.

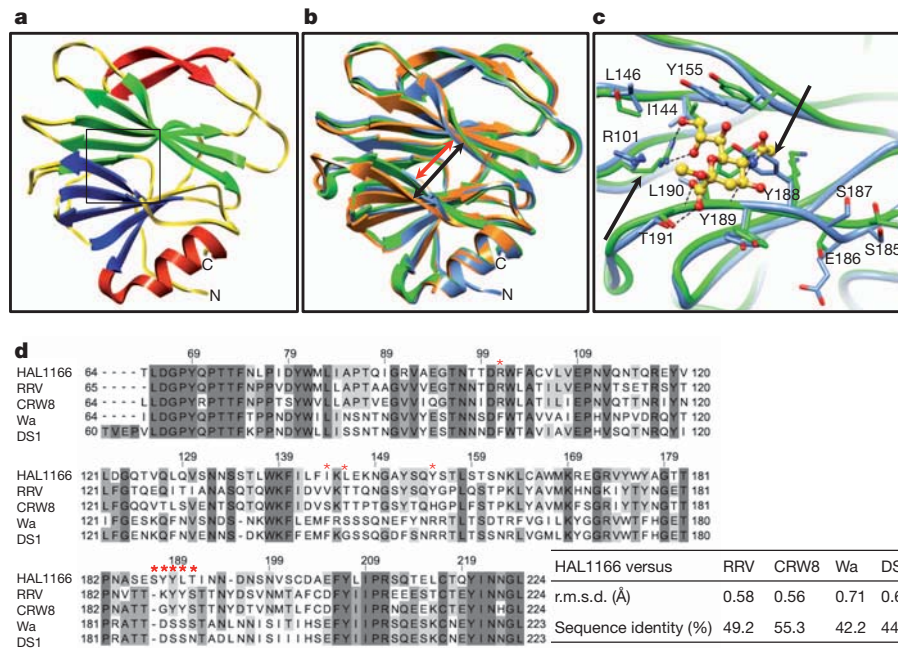


Figure 1 | VP8* structure of HAL1166 P[14] human rotavirus strain and structural comparison with other VP8* structures. **a**, Cartoon representation of the P[14] VP8* structure showing a galectin-like fold with the characteristic two twisted β -sheets (in blue and green) separated by a cleft. The Sia binding site as observed in the animal VP8* is shown by a box. The β -ribbon and the carboxy-terminal α -helix of the structure are shown in red. The amino and carboxy termini are denoted. **b**, Structural alignment of P[14] VP8* structure (blue) with VP8* structure of sialidase-sensitive animal rotavirus RRV strain (green, PDB ID 1KQR) and VP8* of sialidase-insensitive human rotavirus strain Wa (orange, PDB ID 2DWR) shown in the same orientation as in **a**. The width of the cleft in the P[14] VP8* is narrower (red arrow) than in the Wa VP8* structure (black arrow). **c**, P[14] VP8* shows changes in the amino acid composition and side-chain orientations in the region corresponding to Sia binding site of RRV VP8*. The amino acid residues interacting with Sia in the RRV VP8* structure are shown as green sticks, and bound Sia is shown as

yellow sticks with oxygen atoms in red. The residues in this region of the P[14] VP8* structure are shown as blue sticks. The residue numbering corresponds to HAL1166 VP8*. The key amino acid changes in this region between P[14] and RRV VP8* are indicated by black arrows. Noticeable is how Y188 in the P[14] VP8* structure causes steric hindrance if Sia were to bind in this region. Also noticeable is the change in side-chain orientation of the conserved R101. **d**, Alignment of HAL1166 VP8* with other VP8* sequences from sialidase-sensitive animal (RRV, CRW-8) strains, and sialidase-insensitive human rotavirus (Wa and DS1) strains. The residues that interact with Sia in the animal VP8* as shown in **c** are indicated by red stars on the top of the sequence. Highly conserved (>80%), and moderately conserved (>60%) regions are coloured in dark and lighter grey, respectively. The root mean square deviation (r.m.s.d.) of the matching C α atoms between the P[14] VP8* and other VP8* structures along with percentage of sequence identity are shown in the table on the right.

(Fig. 2a). The HBGA binds near the cleft region at the same location as the Sia in the VP8* structures of the animal rotavirus strains (Fig. 2b and Supplementary Fig. 3). The binding of HBGA, which does not cause any conformational changes in the VP8*, involves a network of both direct and solvent-mediated hydrogen bond interactions, in addition to several stabilizing hydrophobic contacts (Supplementary Fig. 4). The terminal GalNAc and Gal of the HBGA participate in all of the interactions (Fig. 2c). The GalNAc participates in direct hydrogen-bonding interactions involving the side chains of R101 and T191, and hydrophobic interactions involving L190 and T191, whereas Gal participates in hydrogen-bonding interactions with the main chain carbonyl groups of Y189 and S187, and hydrophobic interactions with Y189 and Y188. The proximal sugar moieties project out from the surface of VP8* without making any direct contacts with VP8* (Fig. 2 and Supplementary Fig. 3).

The structure of P[14] VP8* with HBGA shows how subtle amino acid changes result in altered ligand specificity. The only positionally conserved common residue between P[14] and animal VP8* that participates in the ligand interactions is R101. Although the position matches well, its side-chain orientation differs significantly between the two structures (Fig. 1c). Without this change, the side-chain of R101 in the P[14] VP8* structure would cause steric hindrance to the GalNAc moiety of the HBGA. Most other residues that participate in the ligand interactions are upstream of position 187 in the VP8* sequence, where an insertion occurs in the P[14] VP8*. This insertion along with sequence variation alters the configuration of the binding site in the P[14] VP8* to allow specific interaction with HBGA.

Insertion at position 187 causes a localized change which makes the side-chain of Y188 clash with Sia when placed in the structure of P[14] VP8* (Fig. 1c). Following this residue, the polypeptide chain reverts back to the same course as in the animal VP8* structures such that Y189 and T191, which interact with HBGA, are now positionally equivalent to Y188 and S190 which interact with Sia in the animal VP8* (Figs 1c and 2c).

The remarkable overlap of the HBGA binding site in the P[14] VP8* with that of the Sia in the animal VP8* structure strongly suggests that A-type HBGA is a cell attachment factor for P[14] rotavirus strains. To examine the biologic relevance of HBGA binding to P[14] VP8*, virus infectivity assays were performed. Using intestinal HT29 cells isolated from a type A individual, dose-dependent abrogation of HAL1166 infectivity was observed, with a greater than 75% reduction at the highest concentration of anti-A-type HBGA antibody compared to an isotype control antibody (Fig. 3a, b). In contrast, this antibody did not inhibit the sialidase-sensitive SA11 rotavirus strain (Fig. 3a). To ascertain further the specificity of the HAL1166 to A-type HBGA, infectivity assays were performed using parental Chinese hamster ovary (CHO) cells, which do not express any HBGA, and genetically-engineered CHO cells expressing either A- or H-type HBGA. CHO cells expressing type A HBGA showed a large increase in infectivity with HAL1166, but not SA11, compared to parental CHO cells or those expressing type H alone (Fig. 3c). Similarly, low infectivity was observed in Caco-2 cells, isolated from a blood type O individual, compared to HT29 cells that express type A HBGA (data not shown). Specificity to A-type HBGA was further confirmed by performing

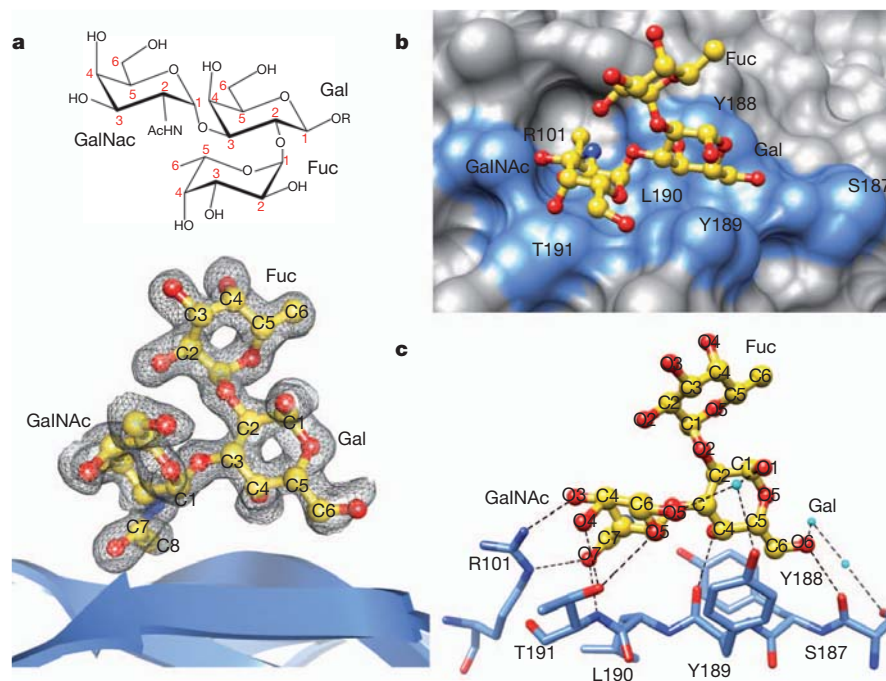


Figure 2 | Structural analysis of P[14] VP8*-A-type HBGA interactions. **a**, The chemical structure of the A-type trisaccharide (above) and simulated annealing omit difference map (below), contoured at 3σ level, showing the binding of A-type trisaccharide to P[14] VP8*. Bound A-type trisaccharide (GalNAc, N-acetylgalactosamine; Gal, galactose; Fuc, fucose) is shown in a ball-and-stick representation (yellow) inside the map with its carbon atoms numbered following the standard convention. The nitrogen and the oxygen atoms in the trisaccharide are coloured in blue and red, respectively. **b**, Surface representation of the P[14] VP8* structure (grey) with the bound A-trisaccharide shown in stick representation (with the same colour scheme as

in **a**). The acetamido group of GalNAc inserts into a well-defined pocket in the VP8* structure. The amino acid residues in the P[14] VP8* which participate in hydrogen bond and hydrophobic interactions with the trisaccharide are indicated in blue. **c**, Network of hydrogen bond interactions (dashed lines) between the VP8* residues (light blue) and A-type trisaccharide (coloured as in **a**). Participating water molecules are shown as small spheres (cyan). More detailed interactions between VP8* and the ligand are given in Supplementary Fig. 4. The terminal two saccharide moieties of the A-type tetrasaccharide (GalNAc α 1-3(Fuc α 1-2)Gal β 1-4GlcNAc) also show similar interactions with the VP8* (Supplementary Fig. 3).

haemagglutination assays using P[14] glutathione-S-transferase (GST)-VP8*. Type A blood cells, but not type O or B, were haemagglutinated by soluble VP8* (Supplementary Fig. 5).

This is the first study describing the structural interactions between a human VP8* and a cellular glycan. Importantly, it shows that binding of sialylated glycans is not obligatory among sialidase-insensitive human rotavirus strains. Our finding that P[14] VP8* specifically recognizes HBGA raises important questions such as whether other

human rotavirus strains interact with similar or other HBGAs in a serotype-dependent manner like in human noroviruses¹⁰, and whether genetically controlled differential expression of HBGAs among world's population plays a role in susceptibility to human rotaviruses. In a recently published paper, G8P[14] rotavirus was identified in the stool samples from two adults with diarrhoea, who lived in the same geographical area in Denmark²¹. The blood type of one of these patients and of another patient infected with a G6P[14] virus was type A (B.

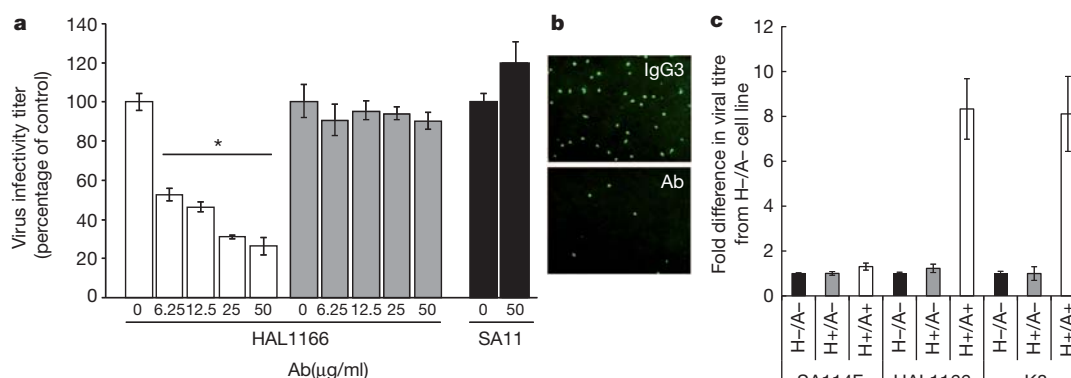


Figure 3 | HAL1166 rotavirus specifically recognizes A-type HBGA. **a**, Dose-dependent inhibition of HAL1166 infection in HT29 cells by anti-A-type antibody (Ab, white bars for HAL1166, and black bars for SA11). Isotype control IgG3 did not inhibit HAL1166 infectivity (grey bars). Error bars (also in Fig. 3c) represent standard deviation and the *P* values were determined by Student's *t*-test, *n* = 3. *All concentrations of anti-A-type antibody reduced infectivity compared to control with *P* < 0.05. **b**, Representative immunofluorescence microscopy images of HT29 cells infected with HAL1166

rotavirus in the presence of $50 \mu\text{g ml}^{-1}$ of IgG3 (top) and anti-A-type antibody (bottom). **c**, Infectivity of SA11 P[1], HAL1166 P[14] and K8 P[9] rotavirus strains in the parental CHO cells (H-/A-), the single transfectant with the Fut2 enzyme (H+/A-), and the double transfectant with both Fut2 and A type histo-blood group glycosyltransferase (H+/A+). The fold difference in infectivity was determined compared to parental cells. For HAL1166 and K8 human rotaviruses, the increase in infectivity in CHO (A+/H+) cells was compared to parental CHO, and CHO (H+/A-), the *P* values were < 0.01.

Böttiger, personal communication). Although this is a small sample size, these findings warrant further epidemiological studies to determine whether HBGA is a susceptibility factor for rotaviruses. Based on sequence comparisons, our prediction that VP8* of the K8 human rotavirus strain (P[9] genotype) would also recognize A-type HBGA (Supplementary Fig. 6) is firmly supported by infectivity assays with parental and derivative CHO cells (Fig. 3c).

The double-stranded RNA rotaviruses, accounting for approximately 500,000 child deaths annually worldwide²², have enormous genetic and strain diversity. In addition to point mutations and gene rearrangements, genetic reassortment between co-circulating strains, similar to influenza viruses, contribute to the expanding diversity of rotaviruses^{23,24}. Current evidence indicates that many of the human rotavirus strains, including the P[14] HAL1166 strain¹⁷, originated from animal reservoirs through reassortment and inter-species transmission^{23,24}. Although effective vaccines are currently available, whether they will remain effective with such expanding virus diversity is an open question^{25,26}. Discovery that a human rotavirus strain with host-switching capabilities binds to a non-sialylated but novel glycan receptor opens new approaches to better understand the molecular basis of critical human rotavirus–host interactions, which probably influences host specificity, cell specificity, pathogenesis and virus evolution.

METHODS SUMMARY

Expression, purification, and crystallization of P[14] VP8* and its complex with A-type oligosaccharides, and structure determination, using RRV VP8* structure (PDB ID: 1KQR) as a molecular replacement model, and refinement was carried out as described in the Methods section. Diffraction data were collected at Baylor College of Medicine using a Rigaku FR-E+ SuperBright rotating anode. See Supplementary Table 2 for data collection and refinement statistics. The carbohydrate-binding specificity of P[14] VP8* was investigated using glycan array v4.2 with 511 glycans in replicates of six (Consortium for Functional Glycomics Protein–Glycan Interaction Core (H) (<http://www.functionalglycomics.com>). GST-tagged VP8* bound on the glycan array was detected using a fluorescent-labelled anti-GST monoclonal antibody. Infectivity assays were performed as previously described^{3,27}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 June 2011; accepted 29 February 2012.

Published online 15 April 2012.

- Blanchard, H., Yu, X., Coulson, B. S. & von Itzstein, M. Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate-binding domain (VP8*). *J. Mol. Biol.* **367**, 1215–1226 (2007).
- Dormitzer, P. R. *et al.* Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core. *J. Virol.* **76**, 10512–10517 (2002).
- Haselhorst, T. *et al.* Sialic acid dependence in rotavirus host cell invasion. *Nature Chem. Biol.* **5**, 91–93 (2009).
- Lopez, S. & Arias, C. F. Early steps in rotavirus cell entry. *Curr. Top. Microbiol. Immunol.* **309**, 39–66 (2006).
- Settembre, E. C., Chen, J. Z., Dormitzer, P. R., Grigorieff, N. & Harrison, S. C. Atomic model of an infectious rotavirus particle. *EMBO J.* **30**, 408–416 (2011).
- Dormitzer, P. R., Sun, Z. Y., Wagner, G. & Harrison, S. C. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21**, 885–897 (2002).
- Kraschnefski, M. J. *et al.* Effects on sialic acid recognition of amino acid mutations in the carbohydrate-binding cleft of the rotavirus spike protein. *Glycobiology* **19**, 194–200 (2009).
- Marionneau, S. *et al.* ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* **83**, 565–573 (2001).

- Ilver, D. *et al.* *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**, 373–377 (1998).
- Glass, R. I., Parashar, U. D. & Estes, M. K. Norovirus gastroenteritis. *N. Engl. J. Med.* **361**, 1776–1785 (2009).
- Matthijnsens, J. *et al.* Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch. Virol.* **156**, 1397–1413 (2011).
- Monnier, N. *et al.* High-resolution molecular and antigen structure of the VP8* core of a sialic acid-independent human rotavirus strain. *J. Virol.* **80**, 1513–1523 (2006).
- Gerna, G. *et al.* Identification of a new VP4 serotype of human rotaviruses. *Virology* **200**, 66–71 (1994).
- Ciarlet, M. & Estes, M. K. Human and most animal rotavirus strains do not require the presence of sialic acid on the cell surface for efficient infectivity. *J. Gen. Virol.* **80**, 943–948 (1999).
- Chitambar, S. D., Arora, R., Kolpe, A. B., Yadav, M. M. & Raut, C. G. Molecular characterization of unusual bovine group A rotavirus G8P[14] strains identified in western India: emergence of P[14] genotype. *Vet. Microbiol.* **148**, 384–388 (2011).
- Fukai, K., Saito, T., Inoue, K. & Sato, M. Molecular characterization of novel P[14]G8 bovine group A rotavirus, Sun9, isolated in Japan. *Virus Res.* **105**, 101–106 (2004).
- Matthijnsens, J. *et al.* Are human P[14] rotavirus strains the result of interspecies transmissions from sheep or other ungulates that belong to the mammalian order Artiodactyla? *J. Virol.* **83**, 2917–2929 (2009).
- Byres, E. *et al.* Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature* **456**, 648–652 (2008).
- Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
- Neu, U. *et al.* Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. *Cell Host Microbe* **8**, 309–319 (2010).
- Midgley, S. E., Hjulsgaard, C. K., Larsen, L. E., Falkenhofst, G. & Böttiger, B. Suspected zoonotic transmission of rotavirus group A in Danish adults. *Epidemiol. Infect.* doi:10.1017/S0950268811001981 (27 September 2011).
- Parashar, U. D., Gibson, C. J., Bresse, J. S. & Glass, R. I. Rotavirus and severe childhood diarrhea. *Emerg. Infect. Dis.* **12**, 304–306 (2006).
- Estes, M. K. & Kapikian, A. Z. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 1917–1974 (Lippincott Williams & Wilkins, 2007).
- Gray, J. & Iturriza-Gomara, M. Rotaviruses. *Methods Mol. Biol.* **665**, 325–355 (2011).
- Angel, J., Franco, M. A. & Greenberg, H. B. Rotavirus vaccines: recent developments and future considerations. *Nature Rev. Microbiol.* **5**, 529–539 (2007).
- Gentsch, J. R. *et al.* Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *J. Infect. Dis.* **192** (Suppl 1), S146–S159 (2005).
- Guillon, P. *et al.* Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* **18**, 1085–1093 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the support from NIH grants AI36040 (to B.V.V.P.), AI 080656 and P30 DK56338 (to M.K.E.), GM62116 (to the Consortium for Functional Glycomics), and the Robert Welch foundation (Q1279) to B.V.V.P. We thank R. Atmar and S. Shanker for helpful discussions and BCM X-ray core facility for data collection.

Author Contributions L.H. carried out expression, purification, crystallization, diffraction data collection and structure determination. L.H., S.E.C., R.C. and N.W.C.-P. contributed to virus infectivity assays in HT29, CHO cells and haemagglutination assays and data analyses. D.F.S. contributed to glycan array experiments and analysis. J.L.P. provided parental and genetically modified CHO cells and advice. M.K.E. provided supervision and advice on cell infectivity assays and analysis. L.H. and B.V.V.P. analysed and interpreted the structural data. B.V.V.P. contributed to the overall direction of the project and wrote the manuscript with input from other authors.

Author Information: The coordinates and structure factors for the P[14] VP8* structures are deposited in the Protein Data Bank under accession numbers 4DRR (apo), 4DRV (with A-type trisaccharide) and 4DSO (with A-type tetrasaccharide). Raw glycan array data are available at <http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.V.V.P. (vprasad@bcm.edu).

METHODS

Protein expression and purification. VP8* (amino acids 64–224) of HAL1166 rotavirus strain (P[14] genotype) was cloned into an expression vector pGEX-2T (GE healthcare) with an N-terminal GST tag and a thrombin cleavage site. The recombinant GST-VP8* was expressed in *E. coli* BL21 (DE3) (Novagen) and purified by Glutathione Sepharose 4 Fast Flow (GE healthcare). The GST tag was cleaved by using thrombin before rebinding the protein mixtures onto a Glutathione Sepharose column to remove the GST, leaving Gly-Ser at the N terminus. The VP8* was then filtered and further purified by size-exclusion chromatography on a Superdex-75 (GE healthcare) column with 10 mM Tris, pH 7.4, 100 mM NaCl, 1 mM dithiothreitol (DTT). The concentration of the purified VP8* was determined by measuring absorbance at 280 nm and using an absorption coefficient of $43,010 \text{ M}^{-1} \text{ cm}^{-1}$ calculated using Vector NTI 11 software (Invitrogen).

Crystallization. Crystallization conditions for P[14] VP8* (13.5 mg ml^{-1}) were screened by hanging-drop vapour diffusion using the Mosquito crystallization robot (TTP LabTech) and visualized using Rock Imager (Formulatrix) at 20°C . The crystals from one of the conditions (30% PEG 1500, sodium acetate trihydrate, pH 4.5) were harvested with the screen condition containing 18% glycerol. To obtain crystals of VP8*–HBGA complex, VP8* was co-crystallized with A-type trisaccharide or tetrasaccharide (purchased from Dextra labs), with a 1:52 or 1:46 excess molar ratio of ligand under similar condition as the unliganded P[14] VP8*.

Data collection and processing. Diffraction data for both unliganded and liganded VP8* crystals were collected at Baylor College of Medicine using Rigaku FR-E+ SuperBright rotating anode. These data were processed with DTREK²⁸ or IMOSFLM as implemented in the CCP4 suite²⁹. Space group was confirmed using POINTLESS³⁰. The unliganded and liganded VP8* structures in the $P2_1$ space group, with one molecule in the asymmetric unit, at $\sim 1.5 \text{ \AA}$ resolution were determined. For initial phasing, the RRV VP8* structure (PDB ID 1KQR) was used as a search model for molecular replacement using Phaser³¹. Following automated model building and solvent addition using ARP/wARP³², the structure was refined using PHENIX³³. The oligosaccharide moieties of the HBGA were generated using the SWEET2 package³⁴ of the Glycosciences.de server (<http://www.glycosciences.de>) and modelled into the electron density using COOT³⁵ and validated by computing simulated annealing omit maps using PHENIX³³. The stereochemistry of the oligosaccharides including the allowed conformational angles was checked using the CARP³⁶ package in the Glycosciences.de server. Data collection and refinement statistics are given in Supplementary Table 2. Ligand interactions were analysed using COOT and LIGPLOT³⁷ with donor to acceptor distances between 2.6 \AA and 3.2 \AA for hydrogen-bonding interactions, and C–C distances between 3.4 \AA and 4.5 \AA for hydrophobic interactions. The structural alignments and calculations of r.m.s.d. were carried out using PyMOL (<http://www.pymol.org/>). Figures were prepared using Chimera³⁸.

Glycan array screening. The carbohydrate-binding specificity of HAL1166 VP8* was investigated on glycan array v4.2 comprised of 511 glycans (Consortium for Functional Glycomics, Protein-Glycan Interaction Core-H) (<http://www.functionalglycomics.org>). Recombinant GST-tagged VP8* at decreasing concentrations in binding buffer (20 mM Tris-HCl pH 7.4, 150 mM sodium chloride, 2 mM calcium chloride, 2 mM magnesium chloride, 0.05% Tween 20, 1% BSA) was applied to separate glycan arrays, and bound protein was detected using a fluorescent-labelled anti-GST monoclonal antibody. Summary of the glycan array results is given in Supplementary Table 1. Concentration dependent binding at $20 \mu\text{g ml}^{-1}$ and $2 \mu\text{g ml}^{-1}$ is shown in Supplementary Fig. 1a, b, where the glycans are ranked according to their relative binding strengths (Supplementary Fig. 1c) as described previously³⁹.

Inhibition and infectivity assays. Inhibition assays were performed on HT29 (human intestinal epithelial) cells following previously described protocols⁴⁰. The monoclonal antibody (MAb) against blood group A antigen (BG-2) was purchased from Covance. The isotype control antibody (MG3-35) was purchased from Abcam. HAL1166 virus was grown as previously described⁴¹. Confluent HT29 cell monolayers were grown on 96-well plates. Increasing concentrations of antibodies were allowed to bind to the cells at 4°C for 1 h before 400 fluorescent focus units (FFU) of virus were added per well. Recombinant VP8* protein competition was determined by treating the cells with 62.5 or $31 \mu\text{g ml}^{-1}$ VP8* at 4°C

for 30 min before virus inoculation as earlier. After allowing virus attachment to HT29 cells for 1 h on ice, the inoculum was removed, and the cells were washed with cold DMEM and incubated for 16 h at 37°C in 95% (v/v) air with 5% (v/v) CO_2 . Virus titres in methanol-fixed cell monolayers were determined by staining cells with rabbit anti-rotavirus antibody and Alexa 488-labelled donkey anti-rabbit secondary antibody (Invitrogen). Virus infectivity was expressed as the percentage of focus-forming units in control wells incubated with the same concentrations of bovine serum albumin. Data are given as the mean of three replicates, and the bar indicates the standard deviation. Data were analysed by Student's *t* test (two-tailed test).

For infectivity assays, SA11, HAL1166 or K8 viruses were serially diluted and incubated on confluent monolayers in 96-well plates of HT29, Caco-2 and CHO cells (parental cells or cells expressing type H HBGA, or type H and type A HBGA prepared as previously described⁴²). Following attachment for 1 h at 37°C , the cells were washed, incubated for 16 h and processed as described above for immunofluorescence.

Haemagglutination assay. Pooled human red blood cells were purchased from Immucor. The RBCs were packed via centrifugation at 500g for 10 min and 0.5% suspensions of each RBC type were prepared in 0.85% saline (pH 6.2). GST (negative control) or GST-tagged VP8* were serially diluted via doubling dilution in PBS (0.01 M sodium phosphate, 0.15 M NaCl, pH 5.5; passed through a $0.2 \mu\text{m}$ pore-size filter) on 96-well V-bottom plates (Nunc) in triplicate. The haemagglutination activity of the VP8 dimers was tested by mixing $50 \mu\text{l}$ of the prep (starting dilution of 10 nM) with an equal volume of the RBC suspensions. Recombinant norovirus virus-like particles (Norwalk virus, genogroup GI.1 and Houston virus, genogroup GII.4) were included as positive controls ($5 \mu\text{g ml}^{-1}$ and $10 \mu\text{g ml}^{-1}$ starting dilutions, respectively) because they have well-characterized haemagglutination activity that is known to be mediated by interaction with histo-blood group antigens on the surface of the red blood cells⁴³. The reaction was allowed to proceed for one hour at 4°C before results were recorded. The titer recorded was the highest dilution of sample that prevented the complete sedimentation of red blood cells to the bottom of the well as compared to the negative controls. A commercial blood typing antibody specific for the B antigen was purchased from Immucor and tested at a starting dilution of 1:100 as another positive control.

28. Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr. D* **55**, 1718–1725 (1999).
29. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
30. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
31. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
32. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).
33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
34. Bohne, A., Lang, E. & von der Lieth, C. W. SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics* **15**, 767–768 (1999).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
36. Lütke, T., Frank, M. & von der Lieth, C. W. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res.* **33**, D242–D246 (2005).
37. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
38. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
39. Smith, D. F., Song, X. & Cummings, R. D. Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol.* **480**, 417–444 (2010).
40. Haselhorst, T. et al. Sialic acid dependence in rotavirus host cell invasion. *Nature Chem. Biol.* **5**, 91–93 (2009).
41. Crawford, S. E. et al. Rotavirus viremia and extraintestinal viral infection in the neonatal rat model. *J. Virol.* **80**, 4820–4832 (2006).
42. Guillon, P. et al. Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* **18**, 1085–1093 (2008).
43. Hutson, A. M., Atmar, R. L., Marcus, D. M. & Estes, M. K. Norwalk virus-like particle hemagglutination by binding to H histo-blood group antigens. *J. Virol.* **77**, 405–415 (2003).

Restoration of vision after transplantation of photoreceptors

R. A. Pearson¹, A. C. Barber¹, M. Rizzi¹, C. Hippert¹, T. Xue³, E. L. West¹, Y. Duran¹, A. J. Smith¹, J. Z. Chuang⁴, S. A. Azam¹, U. F. O. Luhmann¹, A. Benucci², C. H. Sung⁴, J. W. Bainbridge¹, M. Carandini², K.-W. Yau³, J. C. Sowden⁵ & R. R. Ali^{1,6}

Cell transplantation is a potential strategy for treating blindness caused by the loss of photoreceptors. Although transplanted rod-precursor cells are able to migrate into the adult retina and differentiate to acquire the specialized morphological features of mature photoreceptor cells¹, the fundamental question remains whether transplantation of photoreceptor cells can actually improve vision. Here we provide evidence of functional rod-mediated vision after photoreceptor transplantation in adult *Gnat1*^{-/-} mice, which lack rod function and are a model of congenital stationary night blindness². We show that transplanted rod precursors form classic triad synaptic connections with second-order bipolar and horizontal cells in the recipient retina. The newly integrated photoreceptor cells are light-responsive with dim-flash kinetics similar to adult wild-type photoreceptors. By using intrinsic imaging under scotopic conditions we demonstrate that visual signals generated by transplanted rods are projected to higher visual areas, including V1. Moreover, these cells are capable of driving optokinetic head tracking and visually guided behaviour in the *Gnat1*^{-/-} mouse under scotopic conditions. Together, these results demonstrate the feasibility of photoreceptor transplantation as a therapeutic strategy for restoring vision after retinal degeneration.

So far there have been no convincing reports of photoreceptor-cell transplantation actually improving the recipient's vision. This may be due to the relatively low numbers of new rod photoreceptors successfully transplanted in previous studies (typically fewer than 1,000 cells)^{1,3–6}. To establish that new rods truly can improve vision, we optimized the rod-transplantation procedure to increase the number of newly integrated photoreceptor cells in wild-type mice. The donor-cell population was rod-photoreceptor precursors identified by their expression of green fluorescent protein (GFP) under control of the promoter for the rod-specific transcription factor, *Nrl*. We obtained maximum integration together with optimal recipient retinal histology after the transplantation of 200,000 fluorescence-activated cell sorted *Nrl-GFP*⁺ rod precursors, taken from postnatal day (P) 4–8 *Nrl-GFP* donor mice, by subretinal injection to both the superior and the inferior retina. This resulted in a 20- to 30-fold increase in the number of integrated rod photoreceptors compared with previous studies, with up to 26,000 new rods within the outer nuclear layer of recipient adult wild-type mice (16,759 ± 1,705 cells; Fig. 1a and Supplementary Fig. 1). Integrated cells were found predominantly around the injection sites, but were distributed over more than 50% of the retinal area. Thus, up to 16% of donor cells integrated into the host outer nuclear layer (see Supplementary Information for estimation of integrated cell number). Supplementary Fig. 1 summarizes how this was achieved.

To test the functionality of the transplanted rod photoreceptors, we selected a murine model in which improvement of rod vision could be assessed definitively; the *Gnat1*^{-/-} mouse lacks rod α -transducin (*Gnat1*), a protein essential for rod phototransduction. It has no rod

function² or behavioural responses to scotopic visual stimuli; it displays, however, near-normal cone histology⁸, function and behavioural responses to photopic visual stimuli (Supplementary Fig. 2 and refs 2, 9) and shows no loss of cones with time⁸. *Nrl-GFP*⁺ rod precursors integrated into the adult *Gnat1*^{-/-} retina in numbers very similar to those observed in wild-type recipients (18,300 ± 1474; maximum 32,015 integrated cells; Fig. 1a). Integrated cells were appropriately located within the outer nuclear layer and morphologically very similar to wild-type rods (Fig. 1b–i and Supplementary Fig. 3), correctly expressing rod α -transducin (Fig. 1c, d), which is absent in endogenous *Gnat1*^{-/-} rods, together with all other mature rod markers examined (rod arrestin (Fig. 1e, f), rhodopsin, recoverin, phosducin (data not shown)). They demonstrated outer segment formation (Fig. 1b–f) and appropriate light-dependent translocation of rod α -transducin (Fig. 1c, d) and arrestin (Fig. 1e, f)¹⁰ between the outer segments and the cell body/synapse. Most (>80%) of *Nrl-GFP*⁺ rods located within the outer nuclear layer displayed synaptic boutons. All those examined by immunohistochemistry (15–20 cells per marker) appropriately expressed the rod ribbon synapse proteins ribeye, bassoon and/or dystrophin¹¹ (Fig. 1g–i). Finally, ultrastructural analysis confirmed that integrated *Nrl-GFP*⁺ rods typically formed the classic triad synapse with endogenous horizontal and bipolar neurons (Fig. 1j, k; synapses observed for 51/55 cells examined; see Supplementary Information).

To test for light responses in *Nrl-GFP*⁺ rods correctly integrated within *Gnat1*^{-/-} recipients, we performed suction-pipette recordings¹² (Fig. 2a, b). Pigment bleached by the GFP-excitation light used during *Nrl-GFP*⁺ cell searching was regenerated with 9-*cis*-retinal before recording (see Supplementary Information). To control for the effects of the bleaching/regeneration procedure on the speed of the dim-flash response kinetics and the sensitivity of wild-type rods (Fig. 2b), we also recorded from non-injected wild-type retinas with and without the excitation-light/pigment-regeneration procedure (Fig. 2a, b). All recorded *Nrl-GFP*⁺ rods with an intact inner segment, cell body and synaptic terminal showed light responses ($n = 9/9$; Fig. 2a, b). Their saturated photocurrent (4.6 ± 0.5 pA) was similar to that of wild-type rods with or without bleaching/regeneration treatment (Fig. 2b; 6.4 ± 0.7 and 6.1 ± 0.8 pA, respectively). Dim-flash response kinetics, indicated by time-to-peak (t_{peak}) and integration time (t_i), were similar between *Nrl-GFP*⁺ and wild-type rods subjected to bleaching/regeneration (see Fig. 2b). Individual *Nrl-GFP*⁺ rods showed variations in sensitivity, with one cell being as sensitive as bleached/regenerated wild-type rods and the other three over a log unit less sensitive (Fig. 2b). This variation may be real, or simply reflect the small sample size (owing to technical challenges in obtaining this parameter in bleached/regenerated *Nrl-GFP*⁺ rods). For each *Nrl-GFP*⁺ rod recorded, several adjacent GFP^{-ve} rods were also tested; none gave detectable responses (Fig. 2c), thus ruling out light signals originating from cones in the *Gnat1*^{-/-} retina.

¹Department of Genetics UCL Institute of Ophthalmology, University College London, 11–43 Bath Street, London, EC1V 9EL, UK. ²Department of Visual Neuroscience, UCL Institute of Ophthalmology, University College London, 11–43 Bath Street, London, EC1V 9EL, UK. ³Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁴Dyson Vision Research Institute, Department of Ophthalmology, Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, New York 10021, USA.

⁵Developmental Biology Unit, UCL Institute of Child Health, University College London, 30 Guilford Street, London, WC1N 1EH, UK. ⁶Molecular Immunology Unit, UCL Institute of Child Health, University College London, 30 Guilford Street, London, WC1N 1EH, UK.

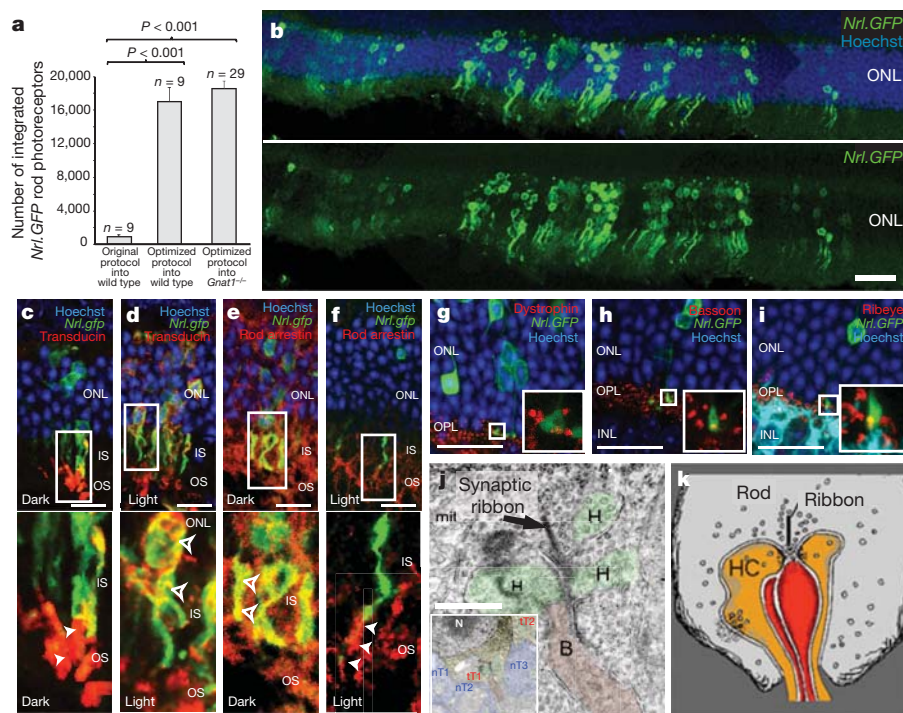


Figure 1 | Improved transplantation protocols significantly improve photoreceptor integration into the adult *Gnat1*^{-/-} model of retinal dysfunction. **a**, *Nrl-GFP*⁺ rod-photoreceptor integration using new and previously published protocols; mean \pm s.e.m., analysis of variance (ANOVA); n , number of eyes. **b**, Typical example of integrated *Nrl-GFP*⁺ rods (green). Scale bar, 50 μ m. **c-f**, Integrated *Nrl-GFP*⁺ rods expressed rod- α -transducin (c, d; red) and rod-arrestin (e, f; red) and demonstrated correct, counter-directional light-mediated translocation of these proteins (bottom panels in c-f, respectively). **g-i**, Integrated *Nrl-GFP*⁺ rods formed spherule synapses and expressed dystrophin (g), bassoon (h) and ribeye (i) (all red). Insets, 3 μ m projections of regions marked. IS, inner segments; OS, outer segments; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer. Scale bar, 10 μ m. **j**, Electron micrographs of low- (inset) and high-power views (consecutive sections), showing *Nrl-GFP*⁺ rod terminals (tT1, tT2). tT1 formed classic triad with horizontal axon terminals (H, orange) and bipolar dendritic terminal (B, red). Endogenous rod terminals (nT; blue) were DAB negative. Arrows, synaptic ribbons; scale bars, 500 nm. **k**, Schematic of rod-triad synapse, reproduced with permission from Webvision (H. Kolb et al., <http://webvision.med.utah.edu/>).

Electroretinography (ERG) provides a gross measure of light-mediated trans-retinal function, averaged across the whole retina. *Gnat1*^{-/-} mice received dual transplants of *Nrl-GFP*⁺ rod precursors into one eye and equivalent sham injections (age-matched *Gnat1*^{-/-} cells or vehicle) or no injection into the contralateral eye (Supplementary Fig. 4). ERGs were recorded weekly 3–6 weeks after transplantation using a double-masked protocol (see Supplementary Information), before the eyes were assessed for *Nrl-GFP*⁺ rod-photoreceptor integration. Despite robust integration (7,135–26,616 integrated cells per eye; $n = 6$), ERG responses were not detected upon scotopic stimulation of procedure *Gnat1*^{-/-} animals (Supplementary Fig. 4a). In contrast,

such responses were readily recordable in wild-type animals (Supplementary Fig. 4f). These results were not unexpected because control experiments with an adeno-associated (AAV2/8) viral vector to deliver a *Gnat1* transgene showed that a scotopic ERG response was recordable only after transduction of approximately 150,000 photoreceptor cells, but not of 60,000 rods (Supplementary Fig. 4d–g). Robust responses were observed in all groups after photopic stimulation with no significant differences between sham and *Nrl-GFP*⁺-treated eyes (Supplementary Fig. 4h). Thus, proper functioning of all retinæ was maintained throughout and was not impaired by the transplantation procedures.

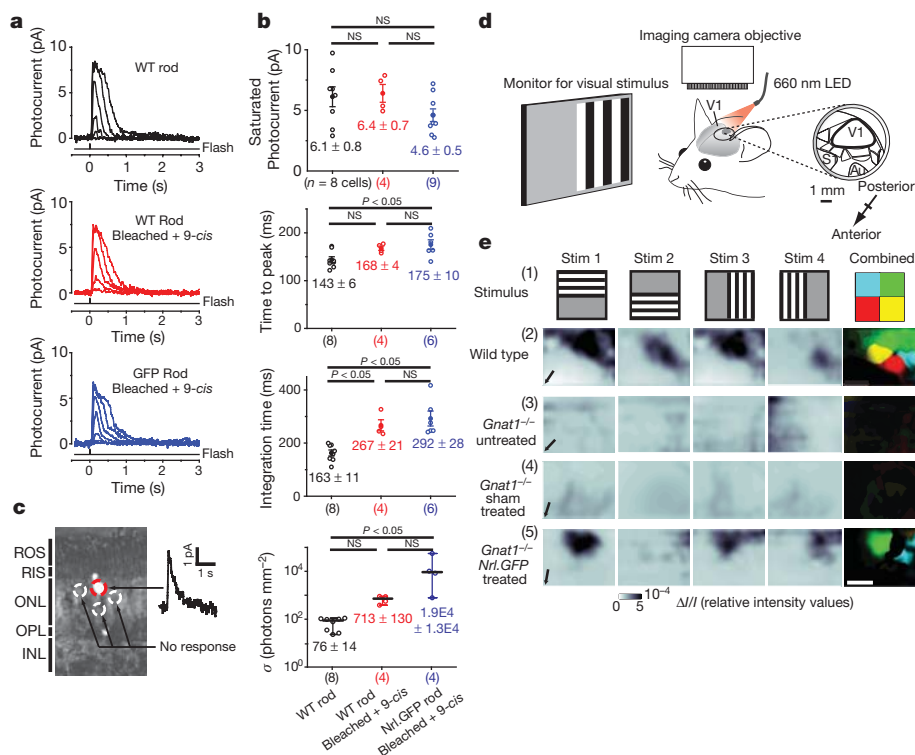


Figure 2 | Transplanted *Nrl-GFP*⁺ rod photoreceptors are light-responsive and project light information to the visual cortex. **a**, Flash-response families from single dark-adapted wild type (black), bleached/regenerated wild type (red) and bleached/regenerated *Nrl-GFP*⁺ (blue) rods. Scale bar, 10 pA. Time (s) is shown on the x-axis. **b**, Comparison of saturated response amplitude and dim-flash response parameters. Individual data points are shown (open circles) with mean \pm s.e.m. (ANOVA, top three panels) or median \pm range (Kruskal-Wallis, bottom panel); n , number of cells; σ , half-saturating flash intensity. Not all parameters were obtainable for each cell. **c**, Representative retinal slice showing light-sensitive *Nrl-GFP*⁺ rod (red circle) and surrounding non-responsive GFP negative rods (white circles). Scale bar, 1 mm. **d**, Schematic of optical intrinsic imaging set-up (see Supplementary Information). Au, auditory cortex; S1, somatosensory cortex. **e**, Visual stimuli (Stim) (1) elicited optical signals in V1 only from *Nrl-GFP*⁺-treated *Gnat1*^{-/-} (5) and wild type (2), but not untreated (3) or sham-injected (4), *Gnat1*^{-/-} eyes. Far right, overlapping parts of the four stimuli were colour-coded (1). Only responses present for two overlapping stimuli were considered genuine sensory-evoked signals and represented with the corresponding colour (2–5). Scale bar, 1 mm.

Given the magnification of retinal signals shown by the visual pathway, we reasoned that photoresponses recorded from transplanted *Nrl-GFP*⁺ rods, although too small in amplitude or too few in number to be detected by ERG, could be transmitted to the brain and generate activity in defined areas of the visual cortex. We therefore performed optical intrinsic imaging^{13,14} of primary visual cortex (Fig. 2d, e and Supplementary Fig. 2b). Four overlapping stimuli covering most of the visual field of the stimulated eye were used and consisted of flickering black/white bars on a grey background (Fig. 2e, row i). When presented to wild-type mice ($n = 5$), such stimuli always evoked strong signals in well-defined areas of the visual cortex under both scotopic (Fig. 2e, row ii) and photopic (Supplementary Fig. 2b) conditions. Normal retinotopy was observed, as indicated by the different cortical locations of the responses to each stimulus (colour maps; Fig. 2e and Supplementary Fig. 2b). In contrast, no detectable response was observed after scotopic stimulation of untreated *Gnat1*^{-/-} mice ($n = 7$) (Fig. 2e, row iii, and Supplementary Figs 2b and 5). We assessed the impact of transplantation by giving *Gnat1*^{-/-} animals *Nrl-GFP*⁺ rod precursors or age-matched *Gnat1*^{-/-} (sham) cells, as before. We observed no cortical activity in response to scotopic visual stimuli

presented to sham-injected eyes ($n = 3$; Fig. 2e, row iv, and Supplementary Fig. 5), demonstrating that neither injection procedure nor presence of non-functional cells led to false signals in the neocortex. However, presentation of the same stimuli to eyes receiving *Nrl-GFP* transplants led to robust responses in all animals ($n = 5$; Fisher's exact test $P = 0.0175$) (Fig. 2e, row v, and Supplementary Fig. 5); partly overlapping stimuli gave responses in likewise partly overlapping areas of the neocortex, with some preservation of retinotopy in the areas of integration (colour maps; Fig. 2e, row v, and Supplementary Fig. 5). Robust integration (10,782–31,075 cells) was observed in all animals displaying cortical activity. Because transplanted cells predominantly integrate near the injection site, it is reasonable that only some stimuli led to a reliable cortical response (compare stimuli 1, 3, 4 with stimulus 2, Fig. 2e, row v). Whenever possible, after using scotopic stimuli we assessed photopic stimuli; such stimuli elicited clear responses in all groups.

To address the paramount question of whether photoreceptor transplantation confers improvements in vision, we assessed visual function first by measuring optomotor head-tracking responses to a rotating grating (Fig. 3a, b; see Supplementary Information)^{9,15,16}.

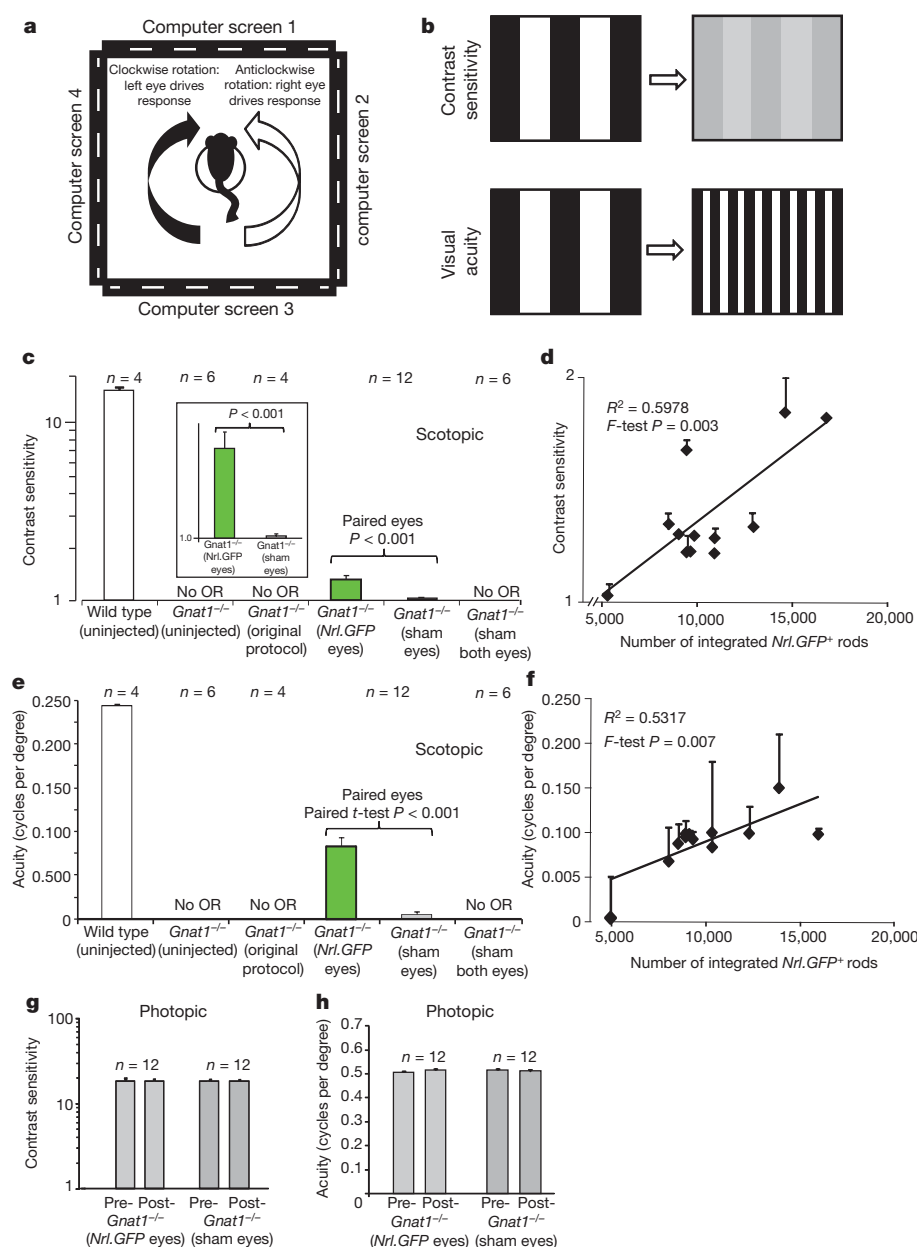


Figure 3 | Rescue of scotopic optokinetic head-tracking behaviour in *Nrl-GFP*-treated *Gnat1*^{-/-} mice. **a**, The Optomotry™ set-up (see Supplementary Information). **b**, Measures of visual function include contrast sensitivity and visual acuity. **c**, **d**, Scotopic contrast sensitivity and visual acuity threshold measurements for *Nrl-GFP*- (green bars) or sham- (*Gnat1*^{-/-}) treated *Gnat1*^{-/-} eyes, and the averages of left and right eyes for *Gnat1*^{-/-} mice receiving original protocol transplants¹ or sham injections to both eyes, or untreated *Gnat1*^{-/-} or wild type (white bars) controls. OR, optomotor response. Paired t -test. **e**, **f**, Scatter plots of contrast sensitivity and visual acuity against integrated *Nrl-GFP*⁺ rod number. **g**, **h**, Photopic contrast sensitivity and visual acuity for *Nrl-GFP*- (light grey) and sham-treated (dark grey) eyes before and after transplantation. Means \pm s.e.m.; n , number of animals.

Gnat1^{-/-} mice received dual transplants of *Nrl-GFP*⁺ rod precursors into one eye and age-matched *Gnat1*^{-/-} cells into the contralateral eye, as previously described. Additional cohorts received (1) the same, but using our original protocol¹, (2) dual *Gnat1*^{-/-} sham injections to both eyes or (3) no injection. As expected, no head-tracking behaviour was observed in any *Gnat1*^{-/-} mouse at scotopic luminance levels (Fig. 3c, e) and normal head tracking occurred under photopic conditions (Supplementary Fig. 2c, d) before transplantation. After transplantation, scotopic optomotor responses were only seen in *Nrl-GFP*-treated eyes; mean contrast-sensitivity threshold for *Nrl-GFP*-treated animals was 1.3 ± 0.1 compared with 15.0 ± 0.6 in wild type (Fig. 3c). Murine visual acuity is poorer under scotopic than photopic conditions (0.245 ± 0.01 versus 0.488 ± 0.003 cycles per degree in wild type; see also ref. 9) but a visual-acuity threshold of 0.097

± 0.007 cycles per degree was recorded in *Nrl-GFP*-treated *Gnat1*^{-/-} mice under the same conditions (Fig. 3e). Moreover, there was a significant positive correlation between both contrast sensitivity (*F*-test, $P < 0.01$; Fig. 3d) and visual acuity (*F*-test, $P < 0.05$; Fig. 3f) and the number of integrated *Nrl-GFP*⁺ rods. Photopic optomotor responses, assessed after scotopic testing, remained unchanged (Fig. 3g, h).

Finally, we assessed vision using the visually guided water-maze task¹⁷, which requires cognitive processing of visual information to associate a grating with escape from a Y-shaped water maze. Under photopic conditions, *Gnat1*^{-/-} mice learned and completed the task as well as wild type ($\geq 70\%$ correct responses; Fig. 4a, Supplementary Fig. 2e–g and Supplementary Information) but performed no better than chance under scotopic conditions. *Gnat1*^{-/-} mice then received (1) dual injections of *Nrl-GFP*⁺ rod precursors to both eyes, (2) sham

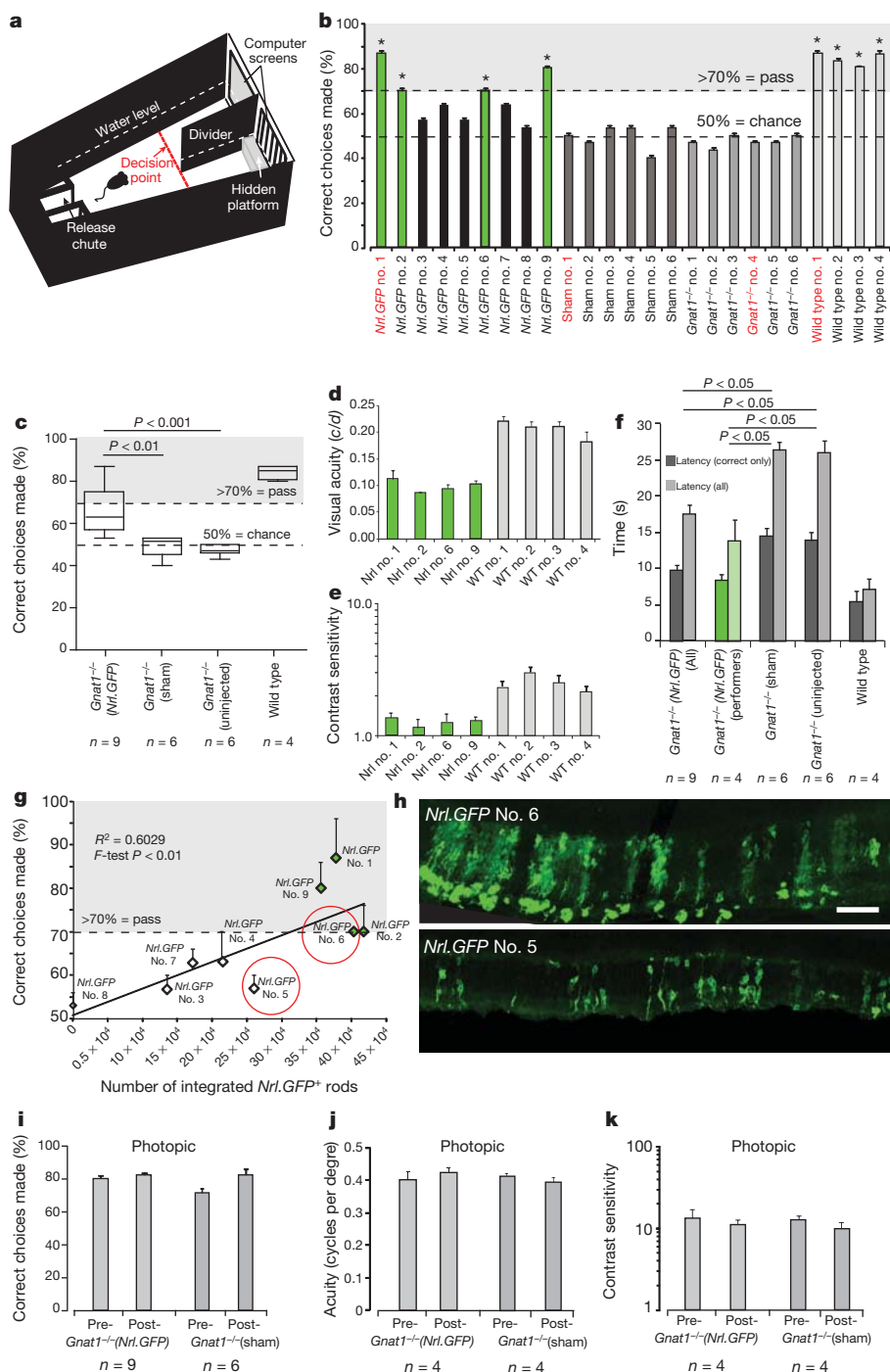


Figure 4 | *Nrl-GFP*-treated *Gnat1*^{-/-} mice can solve the visually guided water-maze task under scotopic conditions. **a**, Schematic of water-maze apparatus (adapted from ref. 22; see Supplementary Information). Mice were trained to associate striped grating with escape from water by a hidden platform. An animal 'passes' a trial by crossing the red line (decision point) on the side of the divider with the striped grating. **b**, Pass rate of *Nrl-GFP*-treated (black), sham-injected (dark grey) and non-injected (mid grey) *Gnat1*^{-/-} and non-injected wild-type (light grey) mice. *Nrl-GFP*-treated animals with a pass-rate of at least 70% are shown in green throughout. Mouse numbers in red refer to mice shown in Supplementary Movie. **c**, Average performance rate of all groups. **d**, Visual acuity and **e**, contrast sensitivity measurements for responders from *Nrl-GFP*-treated (green) and wild-type (light grey) groups. **f**, Swim-time latencies (time-to-platform) for all (light grey) and correct choice-only (dark grey) trials. **g**, Ability to solve water-maze task plotted against integrated *Nrl-GFP* photoreceptor number. **h**, Examples of integration in animals that successfully (top; *Nrl-GFP*-treated, number 6) or unsuccessfully (bottom; *Nrl-GFP*-treated, number 5) solved the task, as indicated in **g** (circled, red). Scale bar, 100 μ m. **i–k**, Pass rate (**i**), visual acuity (**j**) and contrast sensitivity (**k**) for *Nrl-GFP*-treated (light grey bars) and sham-injected (dark grey bars) *Gnat1*^{-/-} mice before and after transplantation under photopic conditions. Means \pm s.e.m.; ANOVA; *n*, number of animals.

injections to both eyes or (3) remained untreated. After transplantation, four out of nine *Gnat1*^{-/-} animals receiving *Nrl-GFP*⁺ rod precursors correctly completed the task in at least 70% of trials under scotopic conditions (green bars, Fig. 4b; Supplementary Movie). In contrast, all sham and untreated *Gnat1*^{-/-} mice performed no better than chance (50%). As a group, *Nrl-GFP*-treated animals were significantly better at solving the water-maze task than either control group (Fig. 4c). The four *Nrl-GFP*-treated *Gnat1*^{-/-} animals able to solve the task correctly further had a contrast sensitivity of 1.3 ± 0.04 and visual acuity of 0.09 ± 0.005 cycles per degree (compared with contrast sensitivity of 3.0 ± 0.24 and visual acuity of 0.206 ± 0.005 cycles per degree, respectively, for wild-type controls; Fig. 4d, e). These same animals solved the task with an average swim time of 15.1 ± 0.8 s (correct choice swim-time latency 9.6 ± 0.8 s), compared with 26.4 ± 1.1 , 26.0 ± 1.2 and 7.1 ± 1.4 s for sham-injected *Gnat1*^{-/-}, untreated *Gnat1*^{-/-} and wild-type mice, respectively (Fig. 4f). There was a significant positive correlation between ability to solve the task and integrated *Nrl-GFP*⁺ rod number (*F*-test, $P < 0.01$; Fig. 4g, h). Cell integration in those animals that performed best in this task was often clustered, rather than widely distributed across the retina, suggesting that proximity of integrated cells to one another may be important, in addition to absolute number. For example, *Nrl-GFP*-treated animals numbers 1 and 2 had similar numbers of integrated rod photoreceptors but number 1 had qualitatively more closely grouped clusters of cells (data not shown). Ability to perform the water-maze task under photopic conditions, assessed after scotopic testing, remained unchanged in all groups after transplantation (Fig. 4i–k).

Although the scotopic visual function recorded here in transplanted *Gnat1*^{-/-} mice is lower than in wild-type animals, such sensitivity is still impressive given that integrated photoreceptors account for less than 1% of the total rods in the retina. However, other studies have demonstrated a remarkable sensitivity from very few light-sensitive cells¹⁸, whereas clinical studies have shown useful vision despite there being no detectable ERG response from the retina¹⁹. Many additional steps are required before these findings can be translated to the clinic, but the results presented here demonstrate for the first time that transplanted rod-photoreceptor precursors can integrate into a dysfunctional adult retina and, by directly connecting with the host retinal circuitry, truly improve vision.

METHODS SUMMARY

The transplantation procedures were performed as described previously^{1,3–5} with modifications, as described in Supplementary Fig. 1 and Supplementary Information. All recipients were adult (6–8 weeks) at time of transplantation and functional assessments were made 4–6 weeks after transplantation. Cellular, retinal and behavioural assessments of visual function were assessed using suction-pipette recordings^{12,20}, ERGs²¹, optical imaging^{13,14}, optomotor^{9,15,16} and water-maze¹⁷ testing and previously published protocols with amendments as detailed in Supplementary Information.

Received 22 June 2011; accepted 28 February 2012.

Published online 18 April 2012.

- MacLaren, R. E. *et al.* Retinal repair by transplantation of photoreceptor precursors. *Nature* **444**, 203–207 (2006).
- Calvert, P. D. *et al.* Phototransduction in transgenic mice after targeted deletion of the rod transducin α -subunit. *Proc. Natl Acad. Sci. USA* **97**, 13913–13918 (2000).
- West, E. L. *et al.* Pharmacological disruption of the outer limiting membrane leads to increased retinal integration of transplanted photoreceptor precursors. *Exp. Eye Res.* **86**, 601–611 (2008).
- Pearson, R. A. *et al.* Targeted disruption of outer limiting membrane junctional proteins (Crb1 and ZO-1) increases integration of transplanted photoreceptor precursors into the adult wild-type and degenerating retina. *Cell Transplant.* **19**, 487–503 (2010).
- West, E. L. *et al.* Long-term survival of photoreceptors transplanted into the adult murine neural retina requires immune modulation. *Stem Cells* **28**, 1997–2007 (2010).
- Bartsch, U. *et al.* Retinal cells integrate into the outer nuclear layer and differentiate into mature photoreceptors after subretinal transplantation into adult mice. *Exp. Eye Res.* **86**, 691–700 (2008).
- Akimoto, M. *et al.* Targeting of GFP to newborn rods by *Nrl* promoter and temporal expression profiling of flow-sorted photoreceptors. *Proc. Natl Acad. Sci. USA* **103**, 3890–3895 (2006).
- Maeda, T. *et al.* A critical role of CaBP4 in the cone synapse. *Invest. Ophthalmol. Vis. Sci.* **46**, 4320–4327 (2005).
- Umino, Y., Solessio, E. & Barlow, R. B. Speed, spatial, and temporal tuning of rod and cone vision in mouse. *J. Neurosci.* **28**, 189–198 (2008).
- Elias, R. V. *et al.* Temporal kinetics of the light/dark translocation and compartmentation of arrestin and α -transducin in mouse photoreceptor cells. *Mol. Vis.* **10**, 672–681 (2004).
- Garner, C. C., Kindler, S. & Gundelfinger, E. D. Molecular determinants of presynaptic active zones. *Curr. Opin. Neurobiol.* **10**, 321–327 (2000).
- Nikonov, S. S. *et al.* Physiological features of the S- and M-cone photoreceptors of wild-type mice from single-cell recordings. *J. Gen. Physiol.* **127**, 359–374 (2006).
- Grinvald, A. *et al.* Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* **324**, 361–364 (1986).
- Schuetz, S., Bonhoeffer, T. & Hubener, M. Mapping retinotopic structure in mouse visual cortex with optical imaging. *J. Neurosci.* **22**, 6549–6559 (2002).
- Alexander, J. J. *et al.* Restoration of cone vision in a mouse model of achromatopsia. *Nature Med.* **13**, 685–687 (2007).
- Prusky, G. T. *et al.* Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. *Invest. Ophthalmol. Vis. Sci.* **45**, 4611–4616 (2004).
- Prusky, G. T., West, P. W. & Douglas, R. M. Behavioral assessment of visual acuity in mice and rats. *Vision Res.* **40**, 2201–2209 (2000).
- Lagali, P. S. *et al.* Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration. *Nature Neurosci.* **11**, 667–675 (2008).
- Berson, E. L. Long-term visual prognoses in patients with retinitis pigmentosa: the Ludwig von Sallmann lecture. *Exp. Eye Res.* **85**, 7–14 (2007).
- Fu, Y. *et al.* Quantal noise from human red cone pigment. *Nature Neurosci.* **11**, 565–571 (2008).
- Tan, M. H. *et al.* Gene therapy for retinitis pigmentosa and Leber congenital amaurosis caused by defects in *AiPL1*: effective rescue of mouse models of partial and complete *AiPL1* deficiency using AAV2/2 and AAV2/8 vectors. *Hum. Mol. Genet.* **18**, 2099–2114 (2009).
- Wong, A. A. & Brown, R. E. Age-related changes in visual acuity, learning and memory in C57BL/6J and DBA/2J mice. *Neurobiol. Aging* **28**, 1577–1593 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Medical Research Council UK (G03000341), the Wellcome Trust (082217), the Royal Society (RG080398), the British Retinitis Pigmentosa Society (GR566) and The Miller's Trust. R.A.P. is a Royal Society University Research Fellow. R.R.A. is partly funded by the Department of Health's National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital and Alcon Research Institute. J.C.S. is supported by Great Ormond Street Hospital Children's Charity. T.X. and K.-W.Y. were supported by a US National Institutes of Health grant (EY06837) and the António Champalimaud Vision Award (Portugal). C.H.S. was supported by grants EY11307, EY016805, Research to Prevent Blindness. M.C. holds the GlaxoSmithKline/Fight for Sight Chair in Visual Neuroscience and is supported by the European Research Council. We thank Y. Umino and the late R. Barlow for advice and training with optomotor recordings, A. Eddoudi for FACS assistance, G. Holder for advice on ERG recordings, C. Hogg for light calibrations, L. Cao for technical suggestions on suction-pipette recordings, and the Department of Genetics, UCL Institute of Ophthalmology, for discussions on the data.

Author Contributions R.A.P. contributed to the concept, design, execution and analysis of all experiments, funding, and wrote the manuscript. A.C.B. contributed to the design, execution and analysis of experiments except single-cell recordings, electron microscopy and behavioural assessments. M.R., A.B. and M.C. contributed to the design, execution and analysis of the intrinsic optical-imaging experiments. C.H., A.J.S. and S.A.A. contributed to the design and execution of the AAV gene supplementation studies, and C.H. the execution of the water-maze tests. Y.D. contributed to the histological processing and execution of the water-maze tests. J.W.B., E.L.W. and U.F.O.L. contributed to the optimization of transplantation protocols. T.X. performed the single-cell recordings. K.W.Y. contributed to the design and analysis of the single-cell recordings. C.H.S. and J.Z.C. performed the ultrastructural analysis. J.C.S. and R.R.A. contributed to the concept and design of the experiments, funding and to manuscript writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.P. (rachael.pearson@ucl.ac.uk) or R.R.A. (r.ali@ucl.ac.uk).

Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia

George T. Eisenhoffer^{1*}, Patrick D. Loftus^{1*}, Masaaki Yoshigi², Hideo Otsuna³, Chi-Bin Chien³, Paul A. Morcos⁴ & Jody Rosenblatt¹

For an epithelium to provide a protective barrier, it must maintain homeostatic cell numbers by matching the number of dividing cells with the number of dying cells. Although compensatory cell division can be triggered by dying cells^{1–3}, it is unknown how cell death might relieve overcrowding due to proliferation. When we trigger apoptosis in epithelia, dying cells are extruded to preserve a functional barrier⁴. Extrusion occurs by cells destined to die signalling to surrounding epithelial cells to contract an actomyosin ring that squeezes the dying cell out^{4–6}. However, it is not clear what drives cell death during normal homeostasis. Here we show in human, canine and zebrafish cells that overcrowding due to proliferation and migration induces extrusion of live cells to control epithelial cell numbers. Extrusion of live cells occurs at sites where the highest crowding occurs *in vivo* and can be induced by experimentally overcrowding monolayers *in vitro*. Like apoptotic cell extrusion, live cell extrusion resulting from overcrowding also requires sphingosine 1-phosphate signalling and Rho-kinase-dependent myosin contraction, but is distinguished by signalling through stretch-activated channels. Moreover, disruption of a stretch-activated channel, *Piezo1*, in zebrafish prevents extrusion and leads to the formation of epithelial cell masses. Our findings reveal that during homeostatic turnover, growth and division of epithelial cells on a confined substratum cause overcrowding that leads to their extrusion and consequent death owing to the loss of survival factors. These results suggest that live cell extrusion could be a tumour-suppressive mechanism that prevents the accumulation of excess epithelial cells.

To examine how cells are eliminated during homeostasis, we immunostained human colon tissues, developing zebrafish epidermises, and cultured Madin–Darby canine kidney (MDCK) epithelial cells for active caspase-3 to identify apoptotic cells, and actin and DNA to highlight cell borders and extrusion^{4,6}. Surprisingly, we found that predominantly caspase-negative cells extrude from the surfaces of colon epithelia (80%, $n = 46$ extruding cells), from zebrafish epidermis (88%, $n = 160$ extruding cells in three experiments), and from overgrown monolayers in culture (67%, $n = 300$ extruding cells in three experiments; Fig. 1a, b, d, e, g, h). By contrast, extrusions resulting from inducing apoptosis in zebrafish with G418 or cultured monolayers with ultraviolet-C were almost exclusively caspase-3 positive⁶. These findings suggest that during homeostasis and development, live rather than dead cells are eliminated by extrusion. Similar live cell extrusions have been observed during mammary gland involution⁷. Because these non-apoptotic *in vivo* extrusions looked identical to apoptotic extrusion, we hypothesized that the extrusion pathway operates in two diverging manners: one to maintain homeostatic cell numbers in epithelia and the other to remove apoptotic/damaged cells.

When quantifying extrusion in both adult human colon and developing zebrafish epidermis, we noticed that extrusions occurred predominantly at the fin edges and colon surfaces, where cell densities were highest (1.7- and 1.8-fold higher than fin centre or crypt side

(Supplementary Fig. 1a, b), respectively, indicated by yellow cells in Fig. 1c, f). Extrusion zones were also more curved *in vivo*; yet, because they occurred most frequently in regions with higher cell density (1.8-fold) compared to proliferative regions in flat cell culture monolayers (Fig. 1i and Supplementary Fig. 1c), we decided to experimentally test if overcrowding strain could induce cells to extrude.

To simulate overcrowding observed *in vivo*, we grew MDCK cells to confluence on a silicone membrane stretched to 28% of its original length and then released it from stretch (Fig. 2a). Within 30 min after release, the number of cells per 100 μm^2 increased 1.3-fold, from 112 ± 5 to 144 ± 4 (Fig. 2b–f). ZO-1 and β -catenin staining confirmed that tight and adherens junctions were still intact and that the average cell diameter had decreased by 30 min after crowding (Fig. 2g–i and Supplementary Fig. 2). Moreover, crowded monolayers maintained adhesion to Cy-5-labelled fibronectin-coated membranes (Fig. 2j). By 6 h after crowding, the number of cells per field equilibrated to pre-release levels (110 ± 10 cells per 100 μm^2 , see Fig. 2e, f), indicating that MDCK epithelia eliminate cells to achieve homeostatic cell numbers.

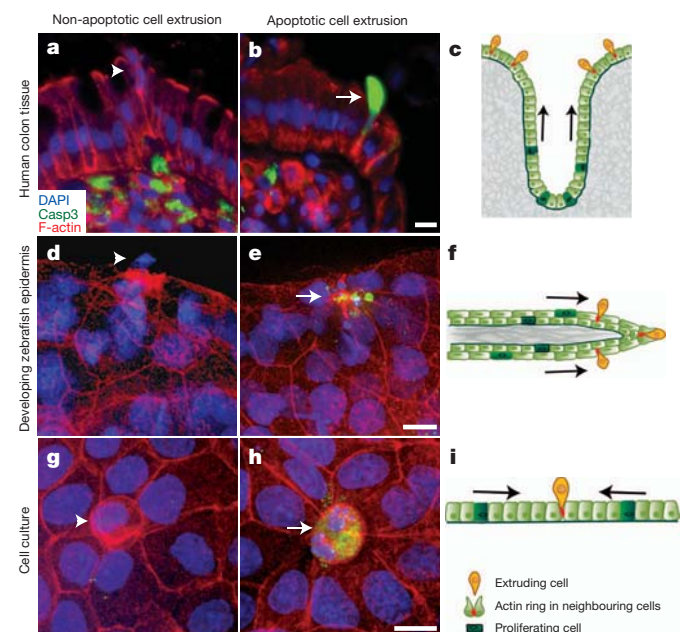


Figure 1 | Both apoptotic and non-apoptotic epithelial cells extrude at locations of high crowding during homeostasis and development. a, b, d, e, g, h, Non-apoptotic (caspase-negative, arrowheads) and apoptotic (caspase-positive, arrows) extrusions at the surface of human colon epithelia (a, b), zebrafish epidermis (d, e), and confluent MDCK monolayers (g, h). c, f, i, Extrusions typically occur at sites of high crowding (indicated by yellow cells). Arrows indicate direction of force from mitoses and migration. Scale bars, 10 μm .

¹Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope Drive, Salt Lake City, Utah 84112, USA. ²Department of Pediatrics, University of Utah, 295 Chipeta Way RM 2S010, Salt Lake City, Utah 84108, USA. ³Department of Neurobiology and Anatomy, University of Utah, 401 MREB, 20N 1900E, Salt Lake City, Utah 84132, USA. ⁴Gene Tools, LLC, 1001 Summerton Way, Philmath, Oregon 97370, USA.

*These authors contributed equally to this work.

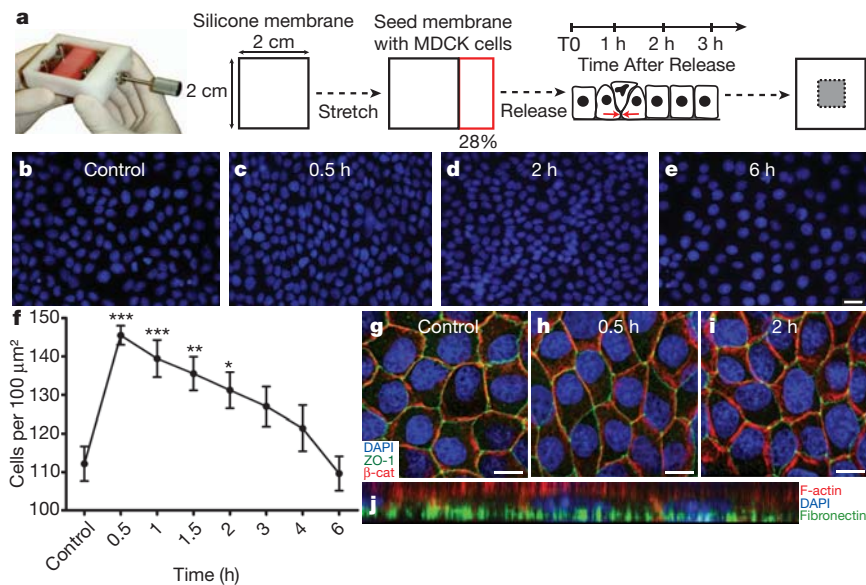


Figure 2 | Characterization of experimental overcrowding. **a**, Experimental design for overcrowding monolayers grown on stretched matrix. Only cells in the grey area (right) were analysed to ensure a homogeneous strain field. **b–f**, Images of the nuclei of control unstretched monolayers and monolayers 0.5–6 h after overcrowding (**b–e**) were quantified in **f**. **f**, Mean number of nuclei per field are plotted, error bars represent s.e.m. **g–i**, Monolayers at 0.5 and 2 h

after release show that adherens and tight junctions maintain integrity after overcrowding, as compared to control. β -cat, β -catenin. **j**, Monolayer stained for actin and DNA at 2 h after release on Cy5-labelled fibronectin shows that contacts to substratum are maintained. Scale bars, 10 μm . *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$.

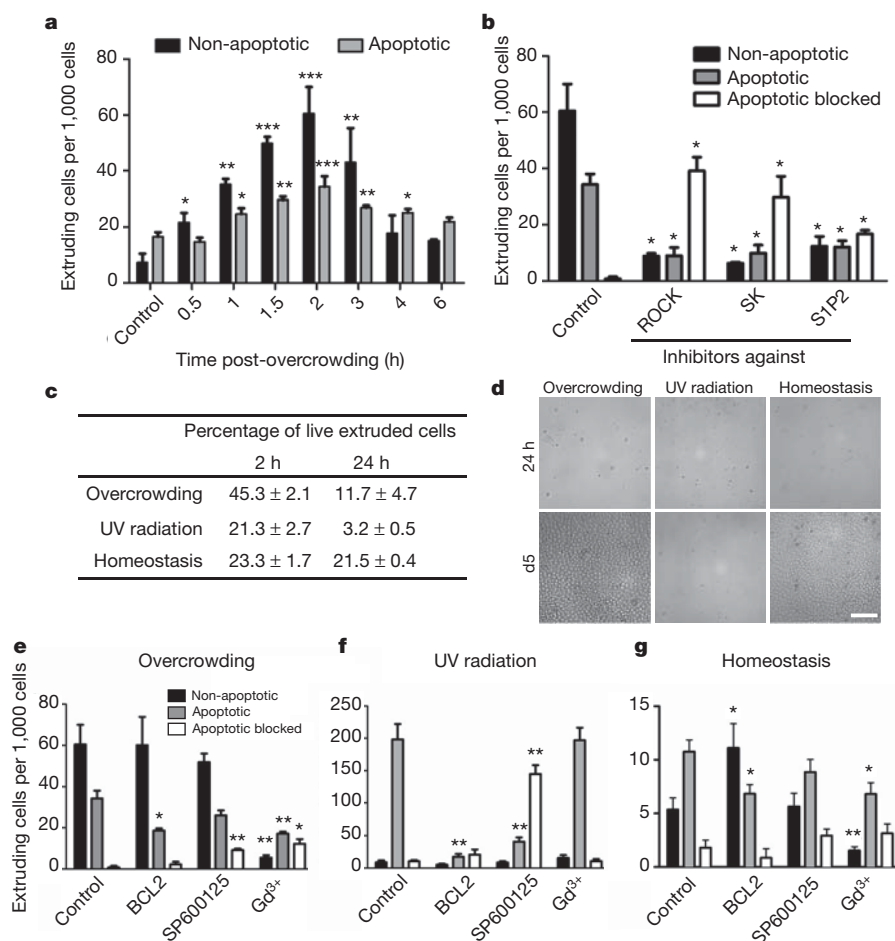


Figure 3 | Overcrowding MDCK monolayers promotes extrusion of non-apoptotic cells. **a**, Quantification of live and apoptotic extruding cells 2 h after overcrowding. **b**, Live cell extrusion requires ROCK, sphingosine kinase (SK) and S1P₂. **c**, Cell viability 2 or 24 h after extrusion. UV, ultraviolet. **d**, Growth of extruded cells collected at 2 h post-crowding after 24 h and 5 days (d5).

e–g, Gd³⁺ inhibits both overcrowding-induced extrusion (**e**) and homeostatic extrusion (**g**) whereas BCL2 overexpression and JNK inhibition only affect ultraviolet-C-induced apoptotic extrusion (**f**). $n \leq 7$ independent experiments, mean \pm s.e.m. Scale bars, 100 μm . *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$.

Immunostaining for actin and active caspase-3 showed that overcrowding induced predominantly non-apoptotic extrusion, as seen during homeostasis (Fig. 3a). Filming crowded MDCK monolayers expressing Lifeact–green fluorescent protein (GFP)⁸ to highlight F-actin, or the active caspase-3 indicator NucView, confirmed that live cells are extruded by contracting actin rings (Supplementary Movie 1 and Supplementary Fig. 4). The numbers of live extruding cells eventually decreased by 6 h after overcrowding, concordantly with the decrease in cell densities. The percentage of non-apoptotic extrusion also correlated to the percentage of crowding (Supplementary Fig. 3).

To determine whether live cell extrusion required the same factors that control apoptotic cell extrusion, we blocked signals required for extrusion during experimental overcrowding. We previously identified sphingosine-1-phosphate (S1P) as the signal apoptotic cells produce to activate the S1P receptor 2 (S1P₂), which triggers Rho-mediated contraction to squeeze the dying cells out⁵. Inhibiting Rho kinase (ROCK) or S1P signalling markedly decreased the percentages of both non-apoptotic and apoptotic cell extrusions (Fig. 3b). Although it was not possible to score non-apoptotic cells blocked from extrusion, as they look like any other cells within the monolayer, the significant decreases in the percentages of non-apoptotic cell extrusion indicated that live cell extrusion, like apoptotic cell extrusion, requires S1P signalling through ROCK-mediated actomyosin contraction.

To assess the fate of overcrowding-induced extruded cells, we analysed their viability using flow cytometry and their ability to survive and proliferate by plating them on new substrata. After collecting extruded cells from the medium 2 h after crowding, 45.3 ± 2.1% were alive, whereas only 21.3 ± 2.7% were alive after ultraviolet-induced apoptotic extrusion (Fig. 3c and Supplementary Fig. 5a). Overcrowding and homeostatically extruded cells were able to proliferate into confluent monolayers, whereas those collected from apoptosis-induced extrusion were not (Fig. 3d). If, however, cells extruded from homeostatic or post-crowded monolayers were collected after 24 h, most cells died instead of proliferating after re-plating (Supplementary Fig. 5b), suggesting that extruded cells typically die unless a new substratum is provided. Presumably, during *in vivo* homeostasis most extruded cells die from anoikis, apoptosis due to loss of survival signals from engagement with the underlying matrix^{9–11}.

We next investigated which signals trigger live cell extrusion during homeostasis or overcrowding. We previously found that overexpression of BCL2 inhibits both apoptosis and extrusion in response to apoptotic stimuli¹² (Fig. 3f); however, BCL2 overexpression did not block live cell extrusion during homeostasis or after overcrowding (Figs 3e, g). To investigate signals that might regulate live cell extrusion after overcrowding, we tested two candidates that are activated by cell stress: the carboxy-terminal JUN kinase (JNK; also known as MAPK8)¹³ and stretch-activated channels^{14,15}. JNK inhibitor blocked apoptotic extrusion in response to ultraviolet-C but not live cell extrusion after overcrowding or homeostatic cell turnover (Figs 3e–g). In contrast, we found that inhibiting stretch-activated ion channels with gadolinium (Gd³⁺)^{16,17} significantly reduced the percentage of both apoptotic and non-apoptotic extrusion events after overcrowding or during epithelial homeostasis, yet had no effect on apoptotic cell extrusion in response to ultraviolet light (Fig. 3e–g). When cell extrusion is blocked by Gd³⁺, the number of cells per 100 μm² remains high at 2 h after crowding (167 ± 12) compared to control crowded cells (112 ± 5), demonstrating that extrusion relieves overcrowded cell densities. Thus, whereas BCL2 and JNK control apoptosis-induced extrusion, stretch-activated signalling controls live cell extrusion during homeostasis that is induced by overcrowding, presumably upstream of S1P signalling (Fig. 4l).

If stretch-activated signalling controls extrusion-mediated homeostatic cell turnover, blocking this pathway *in vivo* should lead to accumulation of excess epithelial cells. By immunostaining and filming the epidermis of developing zebrafish with *Tg(cldnb:lynGFP)*¹⁸, we found that cells proliferate at 37 h post-fertilization (hpf) and migrate

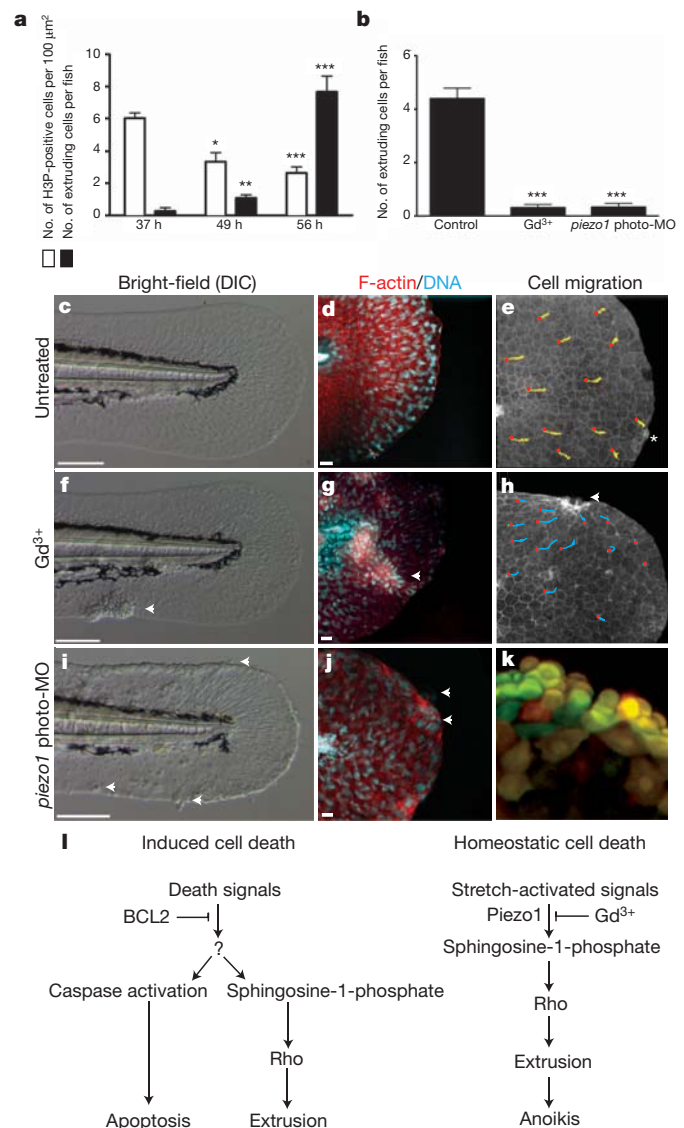


Figure 4 | Disrupting the stretch-activated channel Piezo1 *in vivo* blocks live cell extrusion and causes epithelial mass formation. **a**, Proliferation and extrusion rates during zebrafish tail fin development. H3P, phospho-histone 3. **b**, Gd³⁺ treatment and *piezo1* photo-morpholino (photo-MO) block extrusion ($n = 15–17$ embryos from three independent experiments, mean ± s.e.m.). **c, d, f, g, i, j**, Bright-field and confocal projections of untreated (**c, d**), Gd³⁺-treated (**f, g**), and *piezo1* photo-morpholino knockdown (**i, j**) in 60 hpf developing zebrafish. DIC, differential interference contrast. **e, h**, Tracks from red circle mark the movement of individual epithelial cells in control and Gd³⁺-treated *Tg(cldnb:lynGFP)* developing zebrafish. **k**, Epidermal cell mass from a *piezo1* photo-morpholino knockdown in Kaede zebrafish, where yellow cells are photo-converted. **l**, Model of induced cell death versus homeostatic cell death. Arrowheads denote cell accumulations and asterisk an extrusion. Scale bars in **c, f, i**, 100 μm; in **d, g, j**, 10 μm. *** $P < 0.0005$, ** $P < 0.005$, * $P < 0.05$.

to the fin edges between 48–59 hpf, where they extrude at 56–60 hpf (Fig. 4a, c–e, Supplementary Fig. 6 and Supplementary Movie 2). Treating 32-hpf zebrafish with Gd³⁺ blocked homeostatic cell extrusion and resulted in epidermal cell mass formation at the fin edges, the number of which was dependent on Gd³⁺ concentration (Fig. 4b, f–h, Supplementary Fig. 7 and Supplementary Movie 3). Live imaging of Gd³⁺-treated *Tg(cldnb:lynGFP)* zebrafish revealed that the cell masses formed at zones of convergence, where cells failed to extrude (Fig. 4h and Supplementary Movie 4).

By knocking down candidate stretch-activated channels in zebrafish with morpholino oligonucleotides, we found that live cell extrusion

requires the channel Piezo1 (ref. 14). To prevent developmental defects from early knockdown (Supplementary Fig. 8a, b), we used a photo-cleavable morpholino (photo-morpholino) to knockdown Piezo1 at 30 hpf (Supplementary Fig. 8c, d)¹⁹. By 60 hpf, Piezo1 morphants, like Gd^{3+} -treated fish, could not extrude cells and developed epidermal masses at the fin edges and other sites of tissue strain (Figs 4b, i–k and Supplementary Figs 9 and 10), indicating that Piezo1 regulates extrusion to maintain homeostatic epithelial cell numbers *in vivo*.

As extrusion seems to control cell numbers during homeostasis, we predict that alterations in extrusion could lead to various epithelial pathologies. Aberrant signalling or tensions that block extrusion could lead to formation of cell masses, as those produced in zebrafish when stretch-activated channels are disrupted. Furthermore, these cell masses may set the stage for tumorigenesis by allowing retention of oncogenic or defective cells. In support of this, we found increased cell densities and no clearly identified extrusions in colon polyp sections compared to control sections (Supplementary Fig. 11). Conversely, hyper-contraction of bronchial epithelia following bronchoconstriction in asthmatics could lead to excessive extrusion and characteristic epithelial denuding. The resulting poor barrier could lead to the increased inflammation and infection seen in asthmatics^{20–22}. Whereas mathematical models have suggested that mechanical forces could control tissue homeostasis²³, our results show that cell division and migration within epithelia cause overcrowding strain, which induces live cells to extrude and later die. Thus, homeostatic live cell extrusion provides the previously missing link between proliferation and epithelial cell death (Fig. 1c, f, i).

METHODS SUMMARY

All cell cultures, cell staining, ultraviolet irradiation and imaging were done using MDCK cells, as previously described⁶. Cells were treated with 10 μ M SKI-II (Calbiochem), 10 μ M JTE-013 (Tocris Bioscience), 10 μ M SP600125, 10 μ M Gd^{3+} , (both Sigma-Aldrich), or 1% DMSO as a control. Flow cytometry was done using a Beckman-Dickinson FACScan after treating cells with 1 μ g per 250 μ l of propidium iodide (Sigma-Aldrich). The Huntsman Cancer Institute Tissue Resource and Applications Core provided human colon sections. Developing wild-type AB zebrafish were treated with 10 mM Gd^{3+} at 28.5 °C from 32 hpf until 60 hpf, and immunostained as described previously⁶, or filmed with a Nikon spinning disc confocal microscope using Andor software. The antisense morpholino oligonucleotides and photo-morpholino oligonucleotides were acquired from Gene Tools, LLC. For the photo-morpholino experiments, the translation blocking antisense morpholino was mixed at a 1:1 molar ratio with a 25bp sense photo-morpholino and injected into 1–2-cell-stage wild-type AB or *Et(Gal4-VP16)^{zc1044a};Tg(UAS-E1b:Kaede)^{sl1999t}* zebrafish embryos. At 28–32 hpf, embryos were exposed to 350 nm light for 20 s to release the caging sense morpholino, then fixed and immunostained at 60 hpf.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 April 2011; accepted 27 February 2012.

Published online 15 April 2012.

1. Fan, Y. & Bergmann, A. Apoptosis-induced compensatory proliferation. The Cell is dead. Long live the Cell! *Trends Cell Biol.* **18**, 467–473 (2008).
2. Fan, Y. & Bergmann, A. Distinct mechanisms of apoptosis-induced compensatory proliferation in proliferating and differentiating tissues in the *Drosophila* eye. *Dev. Cell* **14**, 399–410 (2008).
3. Ryoo, H. D., Gorenc, T. & Steller, H. Apoptotic cells can induce compensatory cell proliferation through the JNK and the Wingless signaling pathways. *Dev. Cell* **7**, 491–501 (2004).

4. Rosenblatt, J., Raff, M. C. & Cramer, L. P. An epithelial cell destined for apoptosis signals its neighbors to extrude it by an actin- and myosin-dependent mechanism. *Curr. Biol.* **11**, 1847–1857 (2001).
5. Gu, Y., Forostyan, T., Sabbadini, R. & Rosenblatt, J. Epithelial cell extrusion requires the sphingosine-1-phosphate receptor 2 pathway. *J. Cell Biol.* **193**, 667–676 (2011).
6. Slattum, G., McGee, K. M. & Rosenblatt, J. P115 RhoGEF and microtubules decide the direction apoptotic cells extrude from an epithelium. *J. Cell Biol.* **186**, 693–702 (2009).
7. Monks, J., Smith-Steinhart, C., Kruk, E. R., Fadok, V. A. & Henson, P. M. Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. *Biol. Reprod.* **78**, 586–594 (2008).
8. Riedl, J. *et al.* Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).
9. Frisch, S. M. & Francis, H. Disruption of epithelial cell-matrix interactions induces apoptosis. *J. Cell Biol.* **124**, 619–626 (1994).
10. Gilmore, A. P. Anoikis. *Cell Death Differ.* **12** (suppl. 2), 1473–1477 (2005).
11. Reddig, P. J. & Juliano, R. L. Clinging to life: cell to matrix adhesion and cell survival. *Cancer Metastasis Rev.* **24**, 425–439 (2005).
12. Andrade, D. & Rosenblatt, J. Apoptotic regulation of epithelial cellular extrusion. *Apoptosis* **16**, 491–501 (2011).
13. Tournier, C. *et al.* Requirement of JNK for stress-induced activation of the cytochrome c-mediated death pathway. *Science* **288**, 870–874 (2000).
14. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
15. Olsen, S. M., Stover, J. D. & Nagatomi, J. Examining the role of mechanosensitive ion channels in pressure mechanotransduction in rat bladder urothelial cells. *Ann. Biomed. Eng.* **39**, 688–697 (2011).
16. Bhattacharya, M. R. *et al.* Radial stretch reveals distinct populations of mechanosensitive mammalian somatosensory neurons. *Proc. Natl Acad. Sci. USA* **105**, 20015–20020 (2008).
17. Yang, X. C. & Sachs, F. Block of stretch-activated ion channels in *Xenopus* oocytes by gadolinium and calcium ions. *Science* **243**, 1068–1071 (1989).
18. Haas, P. & Gilmour, D. Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line. *Dev. Cell* **10**, 673–680 (2006).
19. Tallafuss, A. *et al.* Turning gene function ON and OFF using sense- and antisense-photo-morpholinos in zebrafish. *Development*. (in the press).
20. Holgate, S. T. The airway epithelium is central to the pathogenesis of asthma. *Allergol. Int.* **57**, 1–10 (2008).
21. Knight, D. A. & Holgate, S. T. The airway epithelium: structural and functional properties in health and disease. *Respirology* **8**, 432–446 (2003).
22. Swindle, E. J., Collins, J. E. & Davies, D. E. Breakdown in epithelial barrier function in patients with asthma: identification of novel therapeutic approaches. *J. Allergy Clin. Immunol.* **124**, 23–34, quiz 35–36 (2009).
23. Shraiman, B. I. Mechanical feedback as a possible regulator of tissue growth. *Proc. Natl Acad. Sci. USA* **102**, 3318–3323 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Welm, M. Redd and K. Ullman for valuable input on our manuscript and T. Marshall and D. Andrade for Lifeact- and BCL2-expressing MDCK lines. The custom-designed cell culture device was made with the support of National Institute of Biomedical Imaging and Bioengineering Grant EB-4443 to M.Y. (patent pending). This work was supported by a NIH-NIGMS NIH Director's New Innovator Award 1 DP2 OD002056-01 and a Laura and Arthur Colwin Endowed Marine Biology Laboratories Summer Research Fellowship Fund to J.R., and a P30 CA042014 awarded to The Huntsman Cancer Institute for core facilities. An NIH Multidisciplinary Cancer Training Program Grant 5T32 CA03247-8 and American Cancer Society Salt Lake City Postdoctoral Fellowship (120464-PF-11-095-01 CSM) supported G.T.E., the University of Utah Undergraduate Research Opportunities Program Parent Fund Assistantship supported P.D.L., and an NIH R01 Grant MH092256 supported H.O. and C.B.C.

Author Contributions G.T.E. and P.D.L. both contributed to the study design and performed all the experiments. M.Y. designed the stretching apparatus and consulted in the study design. J.R. found preliminary results of live cell extrusion *in vivo* and contributed to the design of the study. G.T.E., J.R. and P.D.L. analysed the data and wrote the paper. P.A.M. designed the photo-cleavable morpholinos, and H.O. and C.-B.C. provided zebrafish with epidermal Kaede expression. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.R. (jody.rosenblatt@hci.utah.edu).

METHODS

Cell culture and overcrowding assays. MDCK II cells were grown as described⁶. A custom-designed Teflon chamber was fabricated to culture cells on flexible silicone membranes in stretched states (2×2 cm, 0.5 mm thickness)^{24,25}. Prior to cell seeding, 2×2 cm silicone membranes were stretched and coated with $5 \mu\text{g ml}^{-1}$ fibronectin (BD Biosciences) at 4°C for 24 h. To stretch membranes with our device, one edge of the silicone membrane was clamped in place, while the other side was clamped to a movable shaft, which moves through a Teflon chamber with a sealing gasket. The movable shaft was pulled out to lengths representing 11%, 22% and 28% strain. 750,000 cells ml^{-1} were plated onto the 2×2 cm silicone matrices in a stretched state in the device or in a culture dish (for non-stretched controls) and grown to confluence, then released from their stretched states, thereby crowding cells in monolayer. Epithelial monolayers were fixed and stained or filmed using a Nikon 90i wide-field fluorescence microscope. BCL2-overexpressing cell lines are described previously¹².

Drug and ultraviolet assays. Confluent MDCK monolayers on silicone membranes were pre-treated with SP600125 JNK inhibitor at 10, 50 and 100 μM or 10, 50 and 100 μM gadolinium III chloride (Gd^{3+}) for 30 min at 37°C . Apoptosis was induced by exposing monolayers to $1,200 \mu\text{J cm}^{-2}$ ultraviolet (254 nm) irradiation in a UV series II (Spectrolite). Cells were fixed and stained from 1–6 h post-ultraviolet treatment depending on the experiment. The minimal drug concentrations (10 μM) established in the ultraviolet-induced extrusion assay were used in the overcrowding and homeostatic assays. For testing extrusion following overcrowding, chambers were pretreated with 10 μM JTE-013 (Tocris Bioscience), 10 μM SKI II (Calbiochem) or 10 μM Y-27632 (Tocris) for 30 min before overcrowding until 2 h after crowding.

Cell immunostaining. Fixation and staining of MDCK cells for actin, DNA, active caspase-3, ZO-1 and β -catenin was carried out as previously described⁶. Fibronectin was labelled with an Amersham Cy5 Ab Labelling Kit (GE Healthcare).

Live cell imaging. Standard MDCK cells or Lifact-GFP-expressing⁸ MDCK cells were grown to confluent monolayers and imaged on a Nikon 90i wide field fluorescent microscope with the stage kept at 37°C . For the NucView experiments, cells were incubated with the activated caspase-3 indicator NucView at 1:200 for 30 min before imaging.

Colon sections. Human colon tissue samples were fixed, imbedded in paraffin, and cut into 10 μm sections by the Huntsman Cancer Institute Tissue Resource and Applications Core. The sections were deparaffinised and rehydrated by incubating in citrus clearing solvent (CCS; Richard Allen Scientific), 100%, 95%, 80%, 70% ethanol, and PBS. Antigens were retrieved by heating the slides in 10 mM sodium citrate at 95°C for 20 min, then rinsed three times with PBS, blocked with 5% BSA/0.5% Tween-20 in PBS for 24 h, and incubated overnight with active caspase-3 antibody, rinsed five times with PBS, incubated in Alexa-488 anti-rabbit antibody, 1 $\mu\text{g ml}^{-1}$ Hoechst, and Alexa-568 anti-actin antibody for 2 h, rinsed three times in PBS, and mounted in Prolong Gold (Invitrogen).

Microscopy. Images were captured on a CTR6000 microscope (Leica) with a $\times 63$ oil lens using a Micromax charge-coupled device camera (Roper Scientific) or on a Nikon Eclipse TE300 inverted microscope converted for spinning disc confocal microscopy (Andor Technologies) using a $\times 20$ or $\times 60$ plan fluor 0.95 oil lens with an electron-multiplied cooled CCD camera $1,000 \times 1,000, 8 \times 8 \text{ mm}^2$ driven by the IQ software (Andor Technologies). Image J was used to stack 8–10 consecutive 1 μm confocal sections into Z series, which were then colour-combined and reconstructed into a three-dimensional image using Metamorph (GE Healthcare). IP Lab 4.0.8 s and Image J software were used to analyse percentages of apoptosis and extrusion. For quantification of extrusion in tissue culture, extruding cells were manually scored as non-apoptotic extruding, apoptotic extruding, or blocked apoptotic extrusions based on the presence of an actin ring and caspase-3 staining in 10,000 monolayer cells.

Statistical analyses. Statistical analysis was done on at least four independent experiments for the overcrowding control and time analysis. Three separate assays were used for drug treatments and for FACS analysis. The error bars in all figures are s.e.m. All *P* values were determined from a two-sided unpaired Student's *t*-test using GraphPad Prism software.

FACS and proliferation analysis. Confluent homeostatic, ultraviolet, and overcrowding monolayers were rinsed three times with PBS to remove any previous cells in suspension. The media was replaced and collected at each denoted analysis time. Media was collected and centrifuged at $1,458g$ in a Damon/IEC Division Clinical Centrifuge for 1 min. The cells were resuspended in 250 μl of PBS containing 4 $\mu\text{g ml}^{-1}$ propidium iodide and analysed by FACS. Samples were analysed on a Beckman Dickinson FACScan and 50–1,000 cells were analysed for propidium iodide fluorescence at each condition. For proliferation analysis, pelleted cells were instead re-plated in a 96-well dish and representative pictures

were taken at 24 h and 5 days after re-plating. Results are from three independent experiments.

Zebrafish care and maintenance. Adult zebrafish were maintained under standard laboratory conditions, with a regular light/dark cycle of 14 h light and 10 h of darkness. Embryos were collected and raised in E3 embryo medium at 28.5°C and staged as previously described²⁶. The Institutional Animal Care and Use Committee (IACUC) at the University of Utah (Animal Welfare Assurance no. 10-07017) have approved all procedures performed in this protocol using the zebrafish *Danio rerio*.

Zebrafish drug treatments. Zebrafish embryos were dechorionated at 24 hpf. At 32 hpf approximately 50–100 embryos were transferred to dishes containing E3 embryo medium with or without (control) 10 mM Gd^{3+} and allowed to develop at 28.5°C until 60 hpf, when the animals were either imaged live or fixed for immunostaining.

Zebrafish immunostaining. Developing zebrafish larvae were fixed in 4% formaldehyde for 1–2 h at room temperature (20°C) or overnight at 4°C . Fixed specimens were then permeabilized by rinses with PBSTx (0.5%) and incubated in blocking buffer (1% DMSO, 2 mg ml^{-1} BSA, 0.5% Triton X-100 and 10% goat serum in PBS) for 2 h. Specimens were then incubated overnight in primary antibodies for phospho-histone H3 (H3P; Abcam, 1:500) or activated caspase-3 (BD Pharmingen, 1:700). Samples were subsequently washed six times with PBSTx and then incubated in blocking buffer for 2 h before incubation with the appropriate secondary antibodies or Alexa-phalloidin (488/568/647) (Invitrogen). After incubation with secondary antibodies or Alexa-phalloidin, the specimens were rinsed five times with PBSTx, incubated with DAPI (1:1,000) for 15 min, rinsed once more and then mounted in Prolong Gold. Depending on the situation, either whole specimens or only tail fragments (from the yolk extension back) were mounted for imaging. For quantification of H3P-positive cells, the region from the cloaca back to the edge of the tail fin was quantified.

Zebrafish imaging. Developing zebrafish were immunostained as described previously⁶ and imaged with either the wide-field fluorescent microscope or a spinning disc confocal microscope, described above. For live imaging, wild-type or Gd^{3+} -treated developing *Tg(cldnb:lynGFP)* zebrafish, a gift from T. Piotrowski, were anaesthetized with 0.02% tricaine in E3, mounted in 1% low melt agarose and imaged on a spinning disc confocal microscope using a $\times 20$ objective, capturing a z-series every 2 min for 3–6 h. Manual tracking of individual cells was performed using Metamorph.

Morpholino antisense oligonucleotide knockdown of Piezo1. All of the morpholino oligonucleotides and photo-morpholino oligonucleotides used in this study were acquired from Gene Tools, LLC. Developing zebrafish embryos were injected with 2–4 ng of a standard translation blocking antisense morpholino oligonucleotide directed against the 5' untranslated region (UTR) of *piezo1* (accession XM_691263) at the 1–2-cell stage and then allowed to develop at 28.5°C . For the photo-cleavable morpholino experiments, the translation blocking antisense morpholino (TBMO) was mixed at a 1:1 molar ratio with a sense-photo morpholino, with a 4 bp mismatch around the photo-linker (see Supplementary Fig. 8c for schematic of strategy) and then injected into 1–2-cell-stage wild-type AB or *Et(Gal4-VP16)^{zc1044a};Tg(UAS-Elb:Kaede)^{s1999t}* embryos. At 28–32 hpf, the developing embryos were then exposed to 350 nm light for 20 s using a $\times 10$ objective on a Nikon 90i wide-field fluorescent microscope to activate the morpholino. Some injected embryos were not converted as a control for 'leakiness' of the sense photo-morpholino. Likewise, some wild-type or Kaede-expressing embryos that were not injected were exposed to 350 nm light to ensure photo-conversion did not cause any adverse effects. Morpholinos: FAM38A 5' UTR antisense morpholino oligonucleotide: GAGCGACACTTCCACTCACATT CCT; FAM38A 5' UTR sense photo-morpholino oligonucleotide: AGGAAT GTGAaxttAGTGTCGCTC.

Western blot analyses of Piezo1. *piezo1* TBMO injected animals ($n = 15$ –20) were collected at 28 hpf and homogenized in lysis buffer, boiled at 85°C for 5 min, spun down and the supernatant collected. A Bradford assay was performed to assess protein concentration. 10–15 μg of protein was run out on a 3–8% Tris-Acetate gel at 150 V for 1 h. Detection was performed with an anti-Piezo1 antibody (Proteintech catalogue no. 15939-1-AP) using standard ECL (enhanced chemiluminescence) methods. HC11, a mammary epithelial cell line, was used as a control for protein levels in mammalian tissues. Piezo1 protein levels are presented relative to the α -tubulin loading control. Quantification of the scanned blot was performed using Image J.

Kaede expression in the zebrafish embryonic epidermis. To generate the *Et(Gal4-VP16;myl7:GFP)^{zc1044a}* enhancer-trap line, a plasmid along with *tol2* mRNA was injected into 1-cell-stage developing zebrafish embryos. The potential founders were then crossed to *Tg(UAS-Elb:Kaede)^{s1999t}* (ref. 27) and subsequently identified by Kaede expression. The identified F1 transgenic embryos were then

imaged at 2 and 5 days post-fertilization for identification and characterization of epithelial-specific expression patterns.

24. Yoshigi, M., Clark, E. B. & Yost, H. J. Quantification of stretch-induced cytoskeletal remodeling in vascular endothelial cells by image processing. *Cytometry A* **55A**, 109–118 (2003).
25. Yoshigi, M., Hoffman, L. M., Jensen, C. C., Yost, H. J. & Beckerle, M. C. Mechanical force mobilizes zyxin from focal adhesions to actin filaments and regulates cytoskeletal reinforcement. *J. Cell Biol.* **171**, 209–215 (2005).
26. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
27. Davison, J. M. *et al.* Transactivation from Gal4-VP16 transgenic insertions for tissue-specific cell labelling and ablation in zebrafish. *Dev. Biol.* **304**, 811–824 (2007).

Osteoprotection by semaphorin 3A

Mikihito Hayashi^{1,2,3}, Tomoki Nakashima^{1,2,3}, Masahiko Taniguchi⁴, Tatsuhiko Kodama⁵, Atsushi Kumanogoh^{6,7}
& Hiroshi Takayanagi^{1,2,3,8}

The bony skeleton is maintained by local factors that regulate bone-forming osteoblasts and bone-resorbing osteoclasts, in addition to hormonal activity. Osteoprotegerin protects bone by inhibiting osteoclastic bone resorption, but no factor has yet been identified as a local determinant of bone mass that regulates both osteoclasts and osteoblasts. Here we show that semaphorin 3A (Sema3A) exerts an osteoprotective effect by both suppressing osteoclastic bone resorption and increasing osteoblastic bone formation. The binding of Sema3A to neuropilin-1 (Nrp1) inhibited receptor activator of nuclear factor- κ B ligand (RANKL)-induced osteoclast differentiation by inhibiting the immunoreceptor tyrosine-based activation motif (ITAM) and RhoA signalling pathways. In addition, Sema3A and Nrp1 binding stimulated osteoblast and inhibited adipocyte differentiation through the canonical Wnt/ β -catenin signalling pathway. The osteopenic phenotype in *Sema3a*^{-/-} mice was recapitulated by mice in which the Sema3A-binding site of Nrp1 had been genetically disrupted. Intravenous Sema3A administration in mice increased bone volume and expedited bone regeneration. Thus, Sema3A is a promising new therapeutic agent in bone and joint diseases.

Bone homeostasis has long been thought to be predominantly maintained by the endocrine system by calcium regulating hormones, but increasing evidence indicates that bone is also under the control of factors related to immune and neuronal regulation^{1,2}. As an imbalance between bone resorption and formation results in metabolic bone disorders such as osteoporosis^{3,4}, understanding the balancing mechanisms is important for the development of therapeutic agents. Because bone formation is linked to resorption through coupling factors^{5,6}, treatment with anti-resorptive agents results in simultaneous suppression of bone formation, with the result that the efficacy is compromised⁶⁻⁹. It is thus crucial to identify a molecule that can regulate both resorption and formation synchronously.

Osteoclasts are derived from monocyte/macrophage precursor cells, and their differentiation is regulated by mesenchymal cells, such as osteoblasts, chondrocytes and osteocytes, which express the key osteoclast differentiation factor RANKL^{1,10-12}. Osteoblastic cells counterbalance the function of RANKL by producing a soluble decoy receptor for RANKL, osteoprotegerin (Opg), the name of which indicates it is a protector of bone¹³. Here we show that conditioned medium from Opg-deficient mouse calvarial cells contains factors that inhibit osteoclast formation, and one of these factors is the axon guidance molecule Sema3A. Sema3A inhibits osteoclast differentiation and promotes osteoblastic bone formation, and is thus a potent osteoprotective factor produced by osteoblastic cells.

Sema3A mediates anti-osteoclastogenesis in osteoblasts

The conditioned medium of osteoblastic cells was able to inhibit osteoclast differentiation of bone marrow-derived monocyte/macrophage precursor cells (BMMs) stimulated by RANKL in the presence of macrophage colony-stimulating factor (M-CSF) (Fig. 1a). We observed a substantial anti-osteoclastogenic effect in the conditioned medium of calvarial cells lacking the *Tnfrsf11b* gene (encoding Opg in mice) (Fig. 1a), suggesting the presence of one or more other soluble

inhibitory factors. To identify the osteoblast-secreted proteins that inhibit osteoclast differentiation, we fractionated the conditioned medium of *Tnfrsf11b*^{-/-} calvarial cells by anion-exchange liquid chromatography (Fig. 1b). We found that fractions 8–10 exerted a potent inhibitory effect on osteoclast differentiation (Fig. 1b). The proteins in fraction 8 were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE; Fig. 1c), and the major bands were excised and analysed by liquid chromatography–tandem mass spectrometry (LC-MS/MS). Among the identified proteins (Fig. 1d), we focused on the axon guidance molecule Sema3A^{14,15} (Supplementary Fig. 1a), as recent studies have suggested that axon guidance molecules are involved in the interaction between osteoblasts and osteoclasts¹⁶⁻¹⁸. Previous reports suggest that Sema3A expressed in the skeletal system may have a role in the regulation of innervation and blood vessel invasion and contribute to skeletal patterning¹⁹⁻²¹, but the function of Sema3A in the regulation of bone remodelling remains unknown.

Western blot analysis confirmed that Sema3A protein was present in fractions 8 and 9 (Supplementary Fig. 1b), and that the inhibitory effect of fraction 9 is largely mediated by Sema3A, as this effect was abrogated by the addition of soluble Nrp1, which functioned as a Sema3A decoy receptor (Supplementary Fig. 1c). The addition of recombinant Sema3A potently inhibited osteoclast differentiation in a dose-dependent manner when Sema3A was added before RANKL treatment (Fig. 1e). When Sema3A was added after RANKL treatment, the inhibitory effect was not observed (Fig. 1e). *Sema3a* was predominantly expressed in osteoblast lineage cells among various cells examined (Supplementary Fig. 2a), whereas Sema3A was not detected in osteoclasts (Supplementary Fig. 2a–c). The expression of *Sema3a* messenger RNA in isolated osteocytes and osteoblasts was comparable (Supplementary Fig. 2d). *Sema3a* mRNA expression in calvarial cells was higher than that of any other semaphorin family member tested (Supplementary Fig. 2e).

¹Department of Cell Signaling, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. ²Japan Science and Technology Agency, Exploratory Research for Advanced Technology Program, Takayanagi Osteonetwork Project, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. ³Global Center of Excellence Program, International Research Center for Molecular Science in Tooth and Bone Diseases, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8549, Japan. ⁴Department of Molecular Medical Sciences, Research Institute for Frontier Medicine, Sapporo Medical University School of Medicine, S-1, W-17, Chuo-ku, Sapporo 060-8556, Japan. ⁵Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, Department of Molecular Biology and Medicine, University of Tokyo, Komaba 4-6-1, Meguro-ku, Tokyo 153-8904, Japan. ⁶Department of Respiratory Medicine, Allergy and Rheumatic Diseases, Graduate School of Medicine, Osaka University, Yamadaoka 2-2, Suita, Osaka 565-0871, Japan. ⁷Department of Immunopathology, Immunology Frontier Research Center, Osaka University, Yamadaoka 3-1, Suita, Osaka 565-0871, Japan. ⁸Centre for Orthopaedic Research, School of Surgery, The University of Western Australia, Nedlands, Western Australia 6009, Australia.

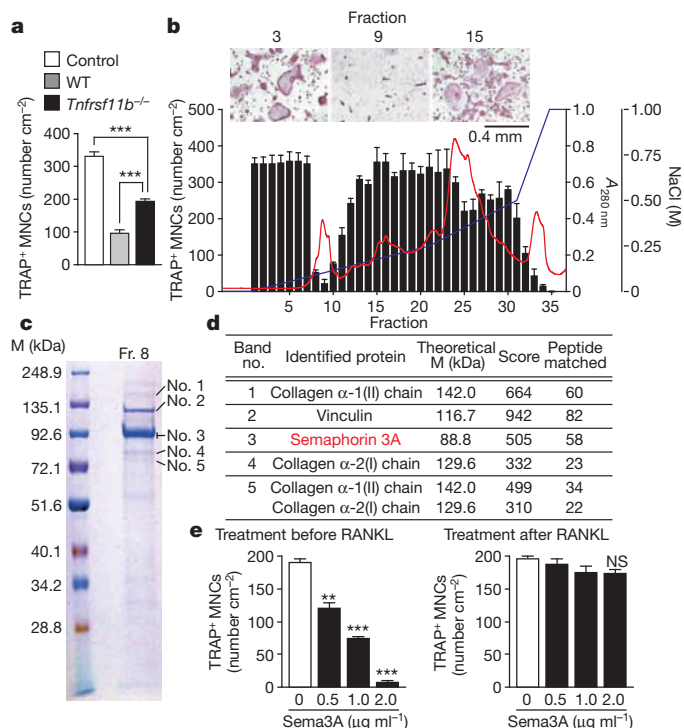


Figure 1 | Identification of Sema3A as an inhibitory factor of osteoclast differentiation. **a**, Effect of wild-type (WT) and Opg-deficient (*Tnfrsf11b*^{-/-}) calvarial cell-conditioned medium on osteoclast differentiation. MNC, mononuclear cells. **b**, Fractionation of Opg-deficient calvarial cell-conditioned medium by anion-exchange chromatography, and the effect of each fraction on osteoclast differentiation (black bars). Absorbance at 280 nm is indicated as a red line and the concentration of NaCl as a blue line. Inset shows TRAP staining of osteoclast cultures treated with fractions 3, 9 and 15. **c**, Coomassie brilliant blue-stained SDS-PAGE image of fraction 8. M, molecular mass. **d**, List of the identified proteins (that is, more than 300 on the MASCOT score), their theoretical molecular mass, MASCOT score and the number of non-redundant peptides. **e**, Effect of Sema3A treatment on osteoclast differentiation. Error bars (**a**, **b** and **e**) denote mean \pm s.e.m. ***P* < 0.01; ****P* < 0.005; NS, not significant.

Sema3A regulates osteoclast differentiation via Nr1p

Microcomputed tomography along with bone morphometric and radiographic analyses showed that *Sema3a*^{-/-} mice¹⁹ have a severe osteopenic phenotype both in trabecular and cortical bones, accompanied by an increase in the osteoclast number and eroded surface (Fig. 2a, b and Supplementary Fig. 3a–c). There was no difference in the number of osteoclast precursor CD11b^{low/-}CD3e⁻B220⁻c-fms⁺c-kit⁺ cells²² in the bone marrow of wild-type and *Sema3a*^{-/-} mice (Supplementary Fig. 3d).

When osteoclast formation was analysed in a coculture of bone marrow and calvarial cells, the formation of tartrate-resistant acid phosphatase (TRAP)-positive multinucleated osteoclasts was markedly enhanced in *Sema3a*^{-/-} cells (Fig. 2c and Supplementary Fig. 3e, f). This enhanced osteoclastogenesis was not observed when BMMs were stimulated by RANKL and M-CSF (Supplementary Fig. 3g). In addition, enhanced osteoclastogenesis was observed in the coculture of wild-type bone marrow cells and *Sema3a*^{-/-} calvarial cells, but not in the coculture of *Sema3a*^{-/-} bone marrow cells and wild-type calvarial cells (Fig. 2c and Supplementary Fig. 3e, f). These results indicate that the osteoblastic expression of Sema3A inhibits osteoclastogenesis. The level of RANKL and Opg in the calvarial cells or the serum was not affected by Sema3A deficiency (Supplementary Fig. 3h, i).

Sema3A binds to a receptor complex of the ligand-binding subunit Nr1p and one of the class A plexins (PlxnA1, PlxnA2, PlxnA3 and PlxnA4), which function as the signal-transducing subunit²³. We found that Nr1p expression in BMMs was rapidly and markedly

suppressed after RANKL stimulation (Fig. 2d). Because Sema3A-induced inhibition of osteoclastogenesis was observed only when Sema3A was added before RANKL stimulation (Fig. 1e), we proposed that Sema3A does not inhibit osteoclastogenesis after RANKL stimulation owing to Nr1p downregulation.

When Nr1p was overexpressed by retroviral transfer, Sema3A exerted an inhibitory effect even when Sema3A had been added after RANKL stimulation (Supplementary Fig. 3j). Notably, osteoclastogenesis was inhibited by Nr1p overexpression only (Supplementary Fig. 3j). When *Nr1p* expression was knocked down by short hairpin RNA (shRNA), the inhibitory effect of Sema3A on osteoclast differentiation was abolished (Supplementary Fig. 3k). Thus, the level of Nr1p correlates with the inhibitory effect of Sema3A on osteoclastogenesis, suggesting that the Nr1p downregulation caused by RANKL signalling is important for proper osteoclast differentiation by cancelling the inhibitory effect of Sema3A.

Furthermore, we analysed knockin mice in which the *Nr1p* gene was replaced by mutant *Nr1p* lacking the Sema-binding site (*Nr1p*^{Sema-} mice), as *Nr1p*^{-/-} mice are embryonically lethal and Nr1p also contains the vascular endothelial growth factor (VEGF)-binding site²⁴. Sema3A did not inhibit RANKL-induced osteoclast differentiation in *Nr1p*^{Sema-} cells (Supplementary Fig. 3l), showing that Sema3A inhibits osteoclastogenesis by binding to Nr1p. As expected, *Nr1p*^{Sema-} mice showed an osteopenic phenotype accompanied by enhanced osteoclast differentiation, which was similar to *Sema3a*^{-/-} mice (Fig. 2e, f and Supplementary Fig. 3m–o).

To understand the mechanism of the RANKL-induced inhibition of Nr1p expression, we examined the involvement of the transcription factors nuclear factor- κ B (NF- κ B), c-Fos and nuclear factor of activated T cells c1 (NFATc1), which are all activated by RANKL¹. RANKL-induced downregulation of *Nr1p* expression was abolished by an NF- κ B inhibitor, but was not affected by the deficiency of NFATc1 or c-Fos (Supplementary Fig. 4a–c). Chromatin immunoprecipitation analysis showed that NF- κ B p65 and, to a lesser extent, p50 were recruited to the proximal NF- κ B-binding site of the *Nr1p* promoter after RANKL stimulation (Supplementary Fig. 4d, e). NF- κ B p65 and p50 inhibited *Nr1p* promoter activity in a reporter gene assay (Supplementary Fig. 4f). Retroviral overexpression of p65 in BMMs led to Nr1p downregulation in the absence of RANKL, which was further facilitated by p50 overexpression (Supplementary Fig. 4g). These effects were dependent on histone deacetylases (Supplementary Fig. 4g, h), suggesting that the recruitment of corepressors²⁵ by RANKL-stimulated NF- κ B is involved in Nr1p downregulation.

Mechanism of anti-osteoclastogenesis by Sema3A

The binding of RANKL to its receptor RANK results in the activation of tumour-necrosis factor (TNF) receptor-associated factor 6 (TRAF6), which stimulates the NF- κ B and mitogen-activated protein kinase (MAPK) pathways¹. RANKL also activates the activator protein 1 (AP-1) transcription factor complex, including c-Fos, which cooperates with NF- κ B to induce NFATc1, thus activating the transcription of osteoclast-specific genes^{1,26}. The robust induction of NFATc1 is dependent on calcium signalling (costimulatory signalling) stimulated by the ITAM-bearing adaptor molecules DNAX-activation protein 12 (DAP12) and Fc receptor common γ subunit (Fc γ R), which associate with immunoglobulin-like receptors such as triggering receptor expressed on myeloid cells 2 (TREM2) and osteoclast-associated receptor (OSCAR)^{1,27}.

From these observations, the question arises as to how the Sema3A–Nr1p axis inhibits osteoclastogenic signalling. RANKL-stimulated induction of the osteoclastic genes *Ctsk*, *Acp5* and *Nfatc1* was severely impaired by Sema3A without affecting the expression of *Tnfrsf11a* or *Csflr* (Fig. 3a). There was no difference in the cell proliferation rate or the percentage of apoptotic cells among the osteoclast precursor cells between the Sema3A-treated and the control cells (Supplementary Fig. 5a, b). RANKL-induced activation

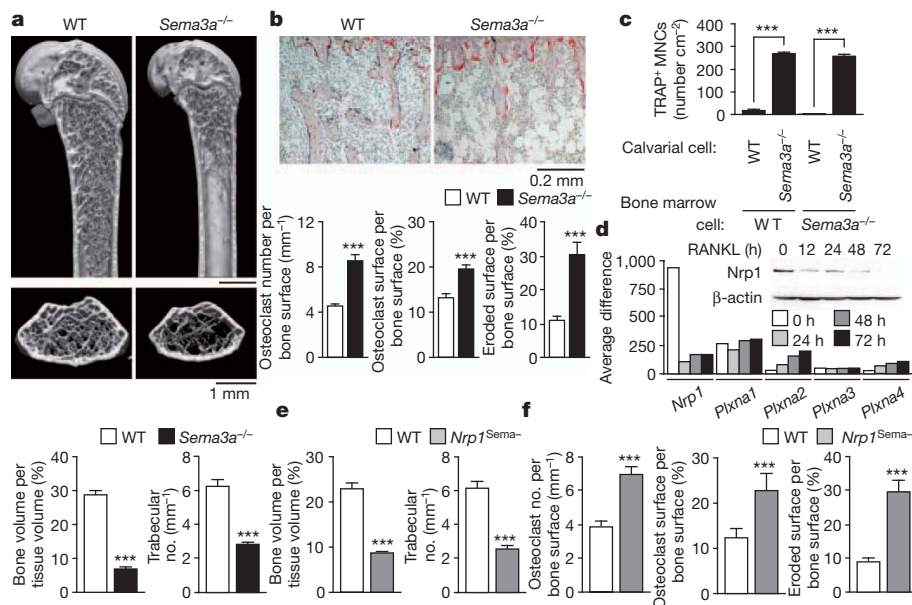


Figure 2 | *Sema3a*^{-/-} and *Nrpl*^{Sema} mice show a severe low bone mass phenotype.

a, Microcomputed tomography images of the femurs of 10-week-old *Sema3a*^{-/-} mice and their wild-type littermates ($n = 4-6$). The bone volume and parameters of trabecular bone were determined by microcomputed tomography analysis. **b**, TRAP staining of the proximal tibiae of *Sema3a*^{-/-} mice and their wild-type littermates ($n = 4-6$). Osteoclastic parameters were measured using bone morphometric analysis. **c**, Osteoclast differentiation from wild-type or *Sema3a*^{-/-} bone marrow cells in coculture with wild-type or *Sema3a*^{-/-} calvarial cells at day 4. **d**, GeneChip analysis of the mRNA expression of *Nrpl* and *PlxnA1-4* during osteoclast differentiation. *Nrpl* protein expression in BMMs stimulated with RANKL was analysed by western blot (inset). **e**, Microcomputed tomography analysis of the femurs of 10-week-old *Nrpl*^{Sema} mice and their wild-type littermates ($n = 4-5$). **f**, Parameters for osteoclastic bone resorption in the bone morphometric analysis of the proximal tibiae of *Nrpl*^{Sema} mice and their wild-type littermates ($n = 4-5$). Error bars (**a-c**, **e** and **f**) denote mean \pm s.e.m. ** $P < 0.01$; *** $P < 0.005$.

of the signalling pathways downstream of TRAF6, including the MAPKs (such as ERK, JNK, p38) and inhibitor of κ B (I κ B) kinases, was comparable in BMMs with or without Sema3A (Supplementary Fig. 5c).

PlxnA1 promotes osteoclast differentiation by activating the ITAM signal through the formation of the PlxnA1-TREM2-DAP12 complex in response to ligands such as Sema6D¹⁷. However, PlxnA1 is constitutively associated with *Nrpl*, which mediates Sema3A signalling

instead of TREM2-DAP12 signalling²⁸. With increasing *Nrpl* expression, the amount of TREM2 associated with PlxnA1 decreased and *Nrpl* associated with PlxnA1 increased (Fig. 3b).

RANKL induced the formation of the PlxnA1-TREM2-DAP12 complex by the downregulation of *Nrpl*, thereby releasing PlxnA1 from the PlxnA1-*Nrpl* complex (Fig. 3c). Sema3A treatment inhibited RANKL-induced formation of the PlxnA1-TREM2-DAP12 complex by inhibiting *Nrpl* downregulation and maintaining the PlxnA1-*Nrpl* complex (Fig. 3c). The cell surface and intracellular expression of *Nrpl* was highly downregulated by RANKL treatment (Supplementary Fig. 5d), but Sema3A treatment induced the internalization of *Nrpl*, as already reported²⁹, and protected RANKL-induced *Nrpl* downregulation without altering *Nrpl* mRNA expression (Supplementary Fig. 5d, e).

RANKL-induced tyrosine phosphorylation of phospholipase C γ 2 (PLC γ 2) and calcium oscillation were both markedly blocked by Sema3A treatment (Supplementary Fig. 5f, g). We observed that osteoclast differentiation in DAP12-deficient (also known as *Tyrbp*-deficient) bone marrow cells was not enhanced even in a coculture with *Sema3a*^{-/-} calvarial cells (Supplementary Fig. 5h), suggesting that Sema3A-induced inhibition is mediated by the modulation of DAP12-induced ITAM signalling. Thus, *Nrpl* competes with TREM2 for PlxnA1, thereby functioning as a suppressor of the PlxnA1-TREM2-DAP12-induced costimulatory signal. Sema3A-induced inhibition of osteoclast differentiation was less observed in the presence of Sema6D (Supplementary Fig. 5i).

We examined the effect of Sema3A on the migration of BMMs because the semaphorin-plexin system regulates actin cytoskeletal rearrangement^{15,23}. We observed a repulsive effect of Sema3A on M-CSF-induced migration of BMMs (Fig. 3d). By contrast, this repulsive effect was not observed in *Nrpl*^{Sema} BMMs (Fig. 3d). Because semaphorin-plexin signalling regulates the Rho family of small GTPases^{15,23}, we examined the effect of Sema3A treatment on M-CSF-induced activation of the RhoA and Rac GTPases. Sema3A treatment abrogated RhoA activation in response to M-CSF (Fig. 3e), but not Rac activation (Supplementary Fig. 5j), suggesting that the inhibition of RhoA activation is involved in the inhibitory effect of Sema3A on the migration of BMMs.

Sema3A regulates osteoblasts through the Wnt pathway

In addition to an osteoclastic phenotype, both *Sema3a*^{-/-} and *Nrpl*^{Sema} mice were found to have osteoblastic and adipocytic

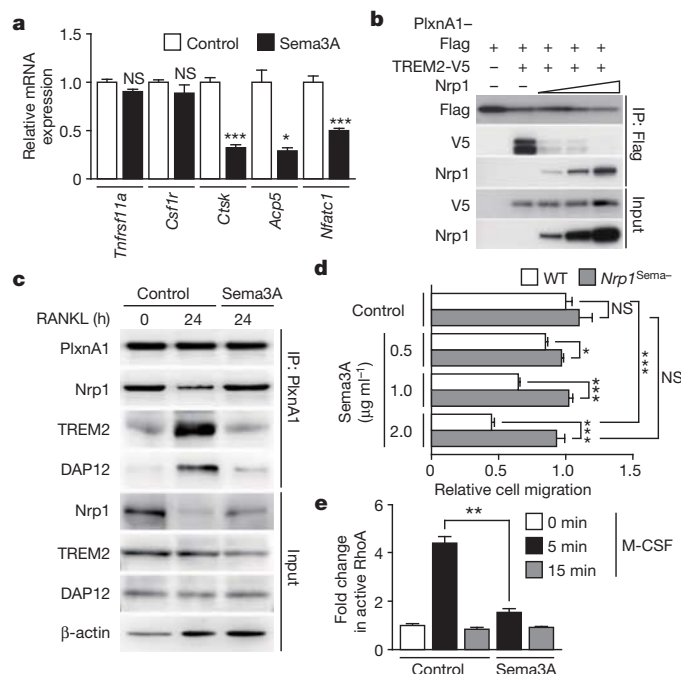


Figure 3 | Inhibition of osteoclast differentiation by Sema3A-Nrpl signalling. **a**, Effect of Sema3A treatment on osteoclastic gene expression in BMMs treated with RANKL for 2 days. **b**, Effect of *Nrpl* expression on the association of PlxnA1 with TREM2. IP, immunoprecipitation. **c**, Effect of Sema3A treatment on the formation of a complex of PlxnA1 with *Nrpl* or TREM2/DAP12 in RANKL-treated BMMs. **d**, Transwell assay of the effect of Sema3A on M-CSF-induced migration of BMMs derived from wild-type or *Nrpl*^{Sema} mice. **e**, Effect of Sema3A treatment on the activation of RhoA in BMMs stimulated with M-CSF. Error bars (**a**, **d** and **e**) denote mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.005$.

phenotypes (Fig. 4a and Supplementary Fig. 6a–c); that is, they had a decreased osteoblast number, a reduced bone formation rate and a markedly increased adipocyte number (Fig. 4b and Supplementary Fig. 6d, e) without any significant difference in the weight of the epididymal white adipose tissue per body weight (Supplementary Fig. 6f, g). Taken together, the severe osteopenic phenotype in *Sema3a*^{−/−} and *Nrp1*^{Sema−} mice was caused by both a decrease in the osteoblastic bone formation and an increase in osteoclastic bone resorption.

Calvarial cells obtained from *Sema3a*^{−/−} or *Nrp1*^{Sema−} mice were cultured in an osteogenic medium with or without Sema3A. Alkaline phosphatase activity and bone nodule formation were markedly decreased in *Sema3a*^{−/−} and *Nrp1*^{Sema−} cells, and Sema3A treatment facilitated the differentiation of *Sema3a*^{−/−} calvarial cells into osteoblastic cells, but not *Nrp1*^{Sema−} cells (Fig. 4c, d and Supplementary Fig. 6h–k). Neither the cell proliferation rate nor the percentage of apoptotic cells was affected in both types of mutant cells (Supplementary Fig. 6l, m). Adipocyte differentiation was highly increased in both *Sema3a*^{−/−} and *Nrp1*^{Sema−} cells, and Sema3A treatment blocked the differentiation of wild-type and *Sema3a*^{−/−} cells into adipocytes, but not *Nrp1*^{Sema−} cells (Fig. 4e and Supplementary Fig. 6n, o). In *Sema3a*^{−/−} cells, the expression of the osteoblast genes *Runx2*, *Sp7* (which encodes osteocalcin), *Alpl* and *Bglap* (encoding osteocalcin) was strongly suppressed (Supplementary Fig. 7a), and the expression of the adipocyte genes *Pparg*, *Cebpa*, *Fabp4* (encoding aP2) and *Lpl* (encoding lipoprotein lipase) was highly increased (Supplementary Fig. 7b). These results indicate that Sema3A activates osteoblast differentiation and inhibits adipocyte differentiation through Nrp1.

Because the mRNA expression levels of the known regulators of mesenchymal cell differentiation^{30,31} were comparable in wild-type and *Sema3a*^{−/−} cells (Supplementary Fig. 7c), we performed gene

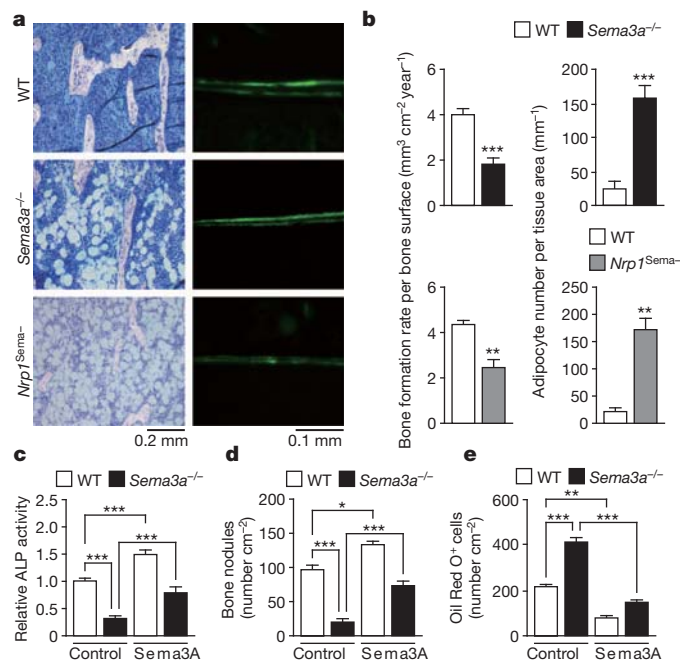


Figure 4 | Impaired osteoblast differentiation and increased adipocyte differentiation in *Sema3a*^{−/−} and *Nrp1*^{Sema−} mice. **a**, Toluidine blue staining of the proximal tibiae of wild-type, *Sema3a*^{−/−} and *Nrp1*^{Sema−} mice (left). New bone formation was determined by calcein double labelling (right). **b**, Osteoblastic and adipocytic parameters measured by histomorphometric analysis of wild-type, *Sema3a*^{−/−} and *Nrp1*^{Sema−} mice (*n* = 4–6). **c**, Alkaline phosphatase (ALP) staining of wild-type and *Sema3a*^{−/−} calvarial cells cultured in osteogenic medium with or without Sema3A. **d**, Bone nodule formation in wild-type and *Sema3a*^{−/−} calvarial cells cultured in osteogenic medium with or without Sema3A. **e**, Adipocyte differentiation in wild-type and *Sema3a*^{−/−} bone marrow stromal cells cultured in adipogenic medium with or without Sema3A. Error bars (b–e) denote mean ± s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.005.

expression profiling of the calvarial cells derived from *Sema3a*^{−/−} mice to obtain insight into the Sema3A-activated molecular pathways in osteoblasts. Gene set enrichment analysis in *Sema3a*^{−/−} cells showed a significant downregulation of the gene sets involved in the Wnt signalling pathway and the Wnt-related signalling pathways (Supplementary Fig. 7d and Supplementary Table 1). We therefore focused on the canonical Wnt pathway, as it is known to promote osteoblast differentiation and inhibit adipocyte differentiation^{31–33}. The mRNA expression of most of the transcriptional targets of β-catenin was considerably reduced (Supplementary Fig. 7e), and the Wnt3a-induced nuclear accumulation of β-catenin was suppressed in *Sema3a*^{−/−} calvarial cells (Fig. 5a).

The Sema3A signalling pathway induces activation of the small G protein Rac1 through FARP2 (FERM, RhoGEF and pleckstrin domain protein 2), which is a guanyl-nucleotide exchange factor (GEF) specific for Rac1 (ref. 34). Previous studies suggested a regulatory role of FARP2 in transduction of semaphorin-induced repulsive cues in axons, via Rac1 activation³⁴. Because Rac1 promotes the nuclear localization of β-catenin in response to canonical Wnt ligands³⁵, we examined the activation of Rac proteins and RhoA in *Sema3a*^{−/−} cells. The activation of Rac, but not RhoA, in *Sema3a*^{−/−} calvarial cells in response to Wnt3a treatment was significantly decreased (Fig. 5b and Supplementary Fig. 7f) as compared with that in wild-type control cells. In addition, Sema3A treatment facilitated the nuclear translocation of β-catenin and the activation of Rac in *Sema3a*^{−/−} cells (Fig. 5a, b). The ectopic expression of a dominant negative form of FARP2 (ΔGEF-FARP2)³⁶ in calvarial cells resulted in the inhibition of osteoblast differentiation, even in the presence of Sema3A (Fig. 5c and Supplementary Fig. 7g). When ΔGEF-FARP2 was overexpressed, reduced nuclear accumulation of β-catenin was observed and Sema3A treatment had no effect on its nuclear localization (Fig. 5d). These results indicate that Sema3A stimulates the canonical Wnt/β-catenin signalling pathway, at least in part, through FARP2-mediated activation of Rac1 during osteoblast differentiation.

Sema3A as an osteoprotective therapeutic agent

To determine the *in vivo* effect of Sema3A administration on bone metabolism, 5-week-old male mice were intravenously injected with recombinant Sema3A or saline once a week. After four weeks of treatment, the trabecular bone volume and trabecular parameters in the distal femur were increased in the Sema3A-treated mice (Fig. 6a and Supplementary Fig. 8a). Bone morphometric analysis showed a decrease in osteoclastic parameters and an increase in osteoblastic parameters (Fig. 6b, c and Supplementary Fig. 8b–e), suggesting that Sema3A exerts a bone-increasing effect by stimulating osteoblastic bone formation and inhibiting osteoclastic bone resorption synchronously. We could not detect any pathological findings in vital organs or any behavioural abnormalities after the Sema3A injection (data not shown). The number of osteoclast precursor cells and osteoprogenitor cells in the bone marrow was not influenced by Sema3A administration (Supplementary Fig. 8f and data not shown), but bone marrow mesenchymal cells derived from Sema3A-treated mice tended to differentiate into osteoblasts instead of adipocytes *in vitro*, although the number of colony forming units was unchanged (Supplementary Fig. 8g, h).

We further investigated the therapeutic potential of Sema3A in a bone regeneration model of cortical bone defects induced by drill hole injury³⁷. Microcomputed tomography analysis showed that the regenerated cortical bone volume in Sema3A-treated mice was higher than in saline-treated mice (Fig. 6d and Supplementary Fig. 8i). The significantly increased osteoblast surface and decreased osteoclast surface around the injured region were observed by histomorphometric analysis (Fig. 6e). These results indicate that the local administration of Sema3A into the injured site accelerates bone regeneration, although we cannot rule out the possibility that Sema3A exerted a bone protective effect partly through the regulation of innervation.

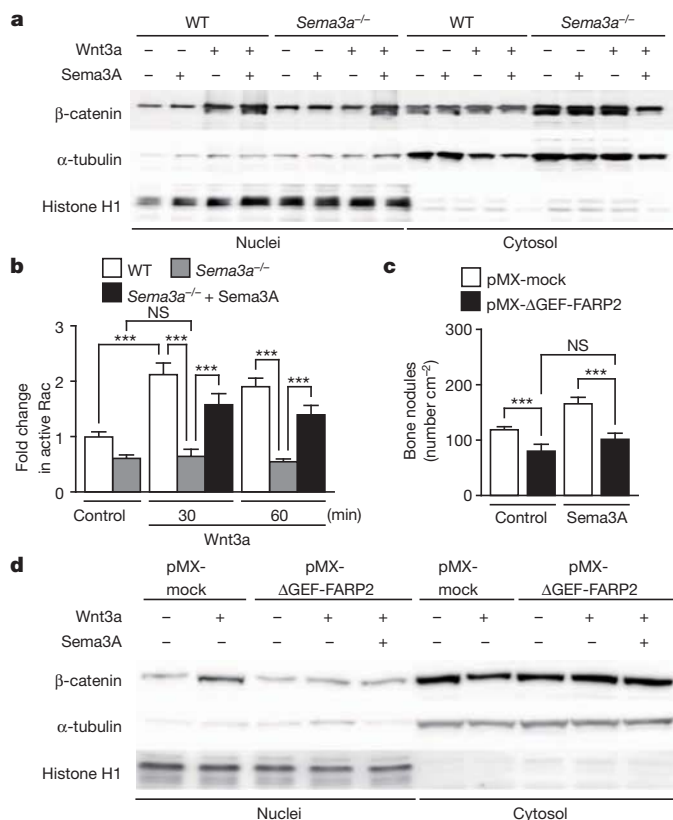


Figure 5 | Regulation of osteoblast differentiation by Sema3A through canonical Wnt signalling. **a**, Analysis of nuclear β -catenin levels in the wild-type and *Sema3A*^{-/-} calvarial cells stimulated by Wnt3a in the presence or absence of Sema3A. Sema3A was simultaneously added with Wnt3a. **b**, Effect of Sema3A treatment on the activation of Rac in the wild-type and *Sema3A*^{-/-} calvarial cells treated with Wnt3a. **c**, Effect of retrovirus-mediated overexpression of Δ GEF-FARP2 on bone nodule formation in calvarial cells cultured in the osteogenic medium in the presence or absence of Sema3A. **d**, Effect of retrovirus-mediated overexpression of Δ GEF-FARP2 on the nuclear localization of β -catenin in calvarial cells treated with Wnt3a in the presence or absence of Sema3A. Error bars (**b**, **c**) denote mean \pm s.e.m. *** P < 0.005.

We examined the effect of Sema3A administration on bone loss in an ovariectomized mouse model of postmenopausal osteoporosis. Ovariectomized 9-week-old mice were treated with a weekly intravenous injection of Sema3A starting two days after ovariectomy and continuing for four weeks. Sema3A administration decreased bone loss after ovariectomy both by inhibiting osteoclastic bone resorption and promoting osteoblastic bone formation (Fig. 6f, g and Supplementary Fig. 8j–l). Recombinant human SEMA3A suppressed osteoclastogenesis and promoted osteoblastogenesis in cultured human cells (Supplementary Fig. 9a, b). These results indicate that Sema3A is a promising potential therapeutic target for bone diseases.

Conclusions

This study demonstrates that the Sema3A expressed by osteoblast lineage cells functions as a potent osteoprotective factor by synchronously inhibiting bone resorption and promoting bone formation (Supplementary Fig. 10a–c). Sema3A represents the long sought soluble molecule with the capacity to bring both osteoblasts and osteoclasts into a condition that favours bone mineral increase. Bone remodelling consists of resorption, transition and formation phases, and the transition phase is under the control of classical coupling factors such as insulin-like growth factor and transforming growth factor- β , which link bone resorption with formation^{38,39}. Sema3A may have a crucial role in the bone formation phase, in which osteoblasts extensively produce bone, and at the same time restrain osteoclasts from

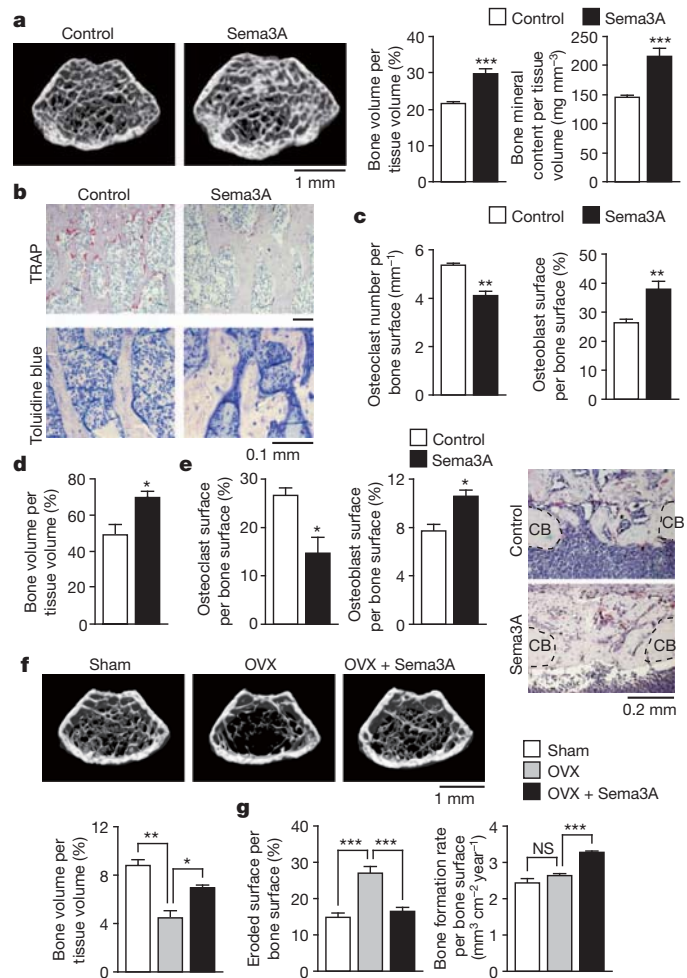


Figure 6 | Sema3A as a potential bone-increasing agent. **a**, Microcomputed tomography analysis of the femurs of 9-week-old wild-type mice treated with Sema3A or saline control (n = 4–7). **b**, Histological analysis of the proximal tibiae of wild-type mice treated with Sema3A or saline control. **c**, Parameters for osteoclasts and osteoblasts in the bone morphometric analysis of wild-type mice treated with Sema3A or saline control (n = 4–7). **d**, Microcomputed tomography analysis of bone regeneration of femoral cortex after drill-hole injury (n = 5). **e**, Histomorphometric analysis of the injured site of the femur (TRAP and haematoxylin staining; n = 5). CB, cortical bone. **f**, Microcomputed tomography analysis of the femurs of the sham-operated (Sham), ovariectomized (OVX) and Sema3A-treated OVX mice (n = 4–5). **g**, Parameters for osteoclastic bone resorption and osteoblastic bone formation in the bone morphometric analysis of the Sham, OVX and Sema3A-treated OVX mice (n = 4–5). Error bars (**a**, **c**–**g**) denote mean \pm s.e.m. * P < 0.05; ** P < 0.01; *** P < 0.005.

migrating to the formation sites and starting to resorb the newly formed bone. The Sema3A protein level in the serum or bone microenvironment could be an auspicious biomarker for bone turnover, as we observed that the serum level of Sema3A decreased with age in mice (data not shown).

The potent anti-osteoclastogenic function of Sema3A is tightly controlled by Nrp1 expression regulated by RANKL signalling. Unless Nrp1 is downregulated by RANKL, the Sema3A–Nrp1 axis inhibits osteoclast differentiation by sequestering PlxnA1 from TREM2 so as to suppress ITAM signalling, and also inhibits RhoA activation via Nrp1–PlxnA to suppress osteoclast precursor cell migration (Supplementary Fig. 10b). After RANKL reduces Nrp1 expression, PlxnA1 associates with TREM2 and DAP12, which facilitate the ITAM-mediated calcium signalling required for osteoclast differentiation (Supplementary Fig. 10b). Because PlxnA1, PlxnA2 and PlxnA3, but not PlxnA4, are expressed by osteoclast and osteoblast

lineage cells (data not shown), the relative contribution of these receptor components should be explored in the future.

Therapeutic agents capable of increasing bone formation have essentially been unavailable except for parathyroid hormone or anti-sclerostin antibody⁴⁰. This study may provide a molecular basis for the development of a combined anti-resorptive and bone-increasing agent capable of facilitating bone regeneration.

METHODS SUMMARY

Mice and bone analysis. The generation of *Sema3a*^{-/-}, *Nrp1*^{Sema-}, *Fos*^{-/-}, *Nfatc1*^{-/-}, *Tnfrsf11b*^{-/-} and *Tyrobp*^{-/-} mice was described previously^{19,24,41–44}. All mice were maintained under specific pathogen-free conditions. All animal experiments were approved by the Institutional Animal Care and Use Committee of Tokyo Medical and Dental University. Three-dimensional micro-computed tomography analyses and bone morphometric analyses were performed as described^{11,14,30,45}. The radiographs were obtained with a high-resolution soft X-ray system (SOFTEX).

Quantitative RT-PCR analysis and GeneChip analysis. Real-time quantitative PCR with reverse transcription (RT-PCR) analysis was performed as described^{11,14,30,45}. In brief, total RNA was extracted by ISOGEN (NIPPON GENE) according to the manufacturer's instructions. First-strand complementary DNAs were synthesized using Superscript III reverse transcriptase (Invitrogen). Quantitative RT-PCR analysis was performed with the LightCycler apparatus (Roche Applied Science) using SYBR Green Realtime PCR Master Mix (TOYOBO). All primer sequences are available on request. GeneChip analysis and gene set enrichment analysis were performed as described previously⁴⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 November 2011; accepted 27 February 2012.

Published online 18 April 2012.

1. Takayanagi, H. Osteoimmunology: shared mechanisms and crosstalk between the immune and bone systems. *Nature Rev. Immunol.* **7**, 292–304 (2007).
2. Elefteriou, F. Regulation of bone remodeling by the central and peripheral nervous system. *Arch. Biochem. Biophys.* **473**, 231–236 (2008).
3. Seeman, E. & Delmas, P. D. Bone quality—the material and structural basis of bone strength and fragility. *N. Engl. J. Med.* **354**, 2250–2261 (2006).
4. Teitelbaum, S. L. & Ross, F. P. Genetic regulation of osteoclast development and function. *Nature Rev. Genet.* **4**, 638–649 (2003).
5. Martin, T. J. & Sims, N. Osteoclast-derived activity in the coupling of bone formation to resorption. *Trends Mol. Med.* **11**, 76–81 (2005).
6. Lewiecki, E. M. New targets for intervention in the treatment of postmenopausal osteoporosis. *Nature Rev. Rheumatol.* **7**, 631–638 (2011).
7. Rachner, T. D., Khosla, S. & Hofbauer, L. C. Osteoporosis: now and the future. *Lancet* **377**, 1276–1287 (2011).
8. Reid, I. R. *et al.* Effects of denosumab on bone histomorphometry: the FREEDOM and STAND studies. *J. Bone Miner. Res.* **25**, 2256–2265 (2010).
9. Odvina, C. V. *et al.* Severely suppressed bone turnover: a potential complication of alendronate therapy. *J. Clin. Endocrinol. Metab.* **90**, 1294–1301 (2005).
10. Suda, T. *et al.* Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr. Rev.* **20**, 345–357 (1999).
11. Nakashima, T. *et al.* Evidence for osteocyte regulation of bone homeostasis through RANKL expression. *Nature Med.* **17**, 1231–1234 (2011).
12. Xiong, J. *et al.* Matrix-embedded cells control osteoclast formation. *Nature Med.* **17**, 1235–1241 (2011).
13. Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
14. Luo, Y., Raible, D. & Raper, J. A. Collapsin: A protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–227 (1993).
15. Tran, T. S., Kolodkin, A. L. & Bharadwaj, R. Semaphorin regulation of cellular morphology. *Annu. Rev. Cell Dev. Biol.* **23**, 263–292 (2007).
16. Negishi-Koga, T. *et al.* Suppression of bone formation by osteoclastic expression of semaphorin 4D. *Nature Med.* **17**, 1473–1480 (2011).
17. Takegahara, N. *et al.* Plexin-A1 and its interaction with DAP12 in immune responses and bone homeostasis. *Nature Cell Biol.* **8**, 615–622 (2006).
18. Matsuo, K. & Irie, N. Osteoclast-osteoblast communication. *Arch. Biochem. Biophys.* **473**, 201–209 (2008).
19. Taniguchi, M. *et al.* Disruption of semaphorin III/D gene causes severe abnormality in peripheral nerve projection. *Neuron* **19**, 519–530 (1997).
20. Gomez, C. *et al.* Expression of Semaphorin-3A and its receptors in endochondral ossification: potential role in skeletal development and innervation. *Dev. Dyn.* **234**, 393–403 (2005).
21. Behar, O., Golden, J. A., Mashimo, H., Schoen, F. J. & Fishman, M. C. Semaphorin III is needed for normal patterning and growth of nerves, bones and heart. *Nature* **383**, 525–528 (1996).
22. Jacquin, C., Gran, D. E., Lee, S. K., Lorenzo, J. A. & Aguila, H. L. Identification of multiple osteoclast precursor populations in murine bone marrow. *J. Bone Miner. Res.* **21**, 67–77 (2006).
23. Neufeld, G. & Kessler, O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nature Rev. Cancer* **8**, 632–645 (2008).
24. Gu, C. *et al.* Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. *Dev. Cell* **5**, 45–57 (2003).
25. Ashburner, B. P., Westerheide, S. D. & Baldwin, A. S. Jr. The p65 (RelA) subunit of NF- κ B interacts with the histone deacetylase (HDAC) corepressors HDAC1 and HDAC2 to negatively regulate gene expression. *Mol. Cell. Biol.* **21**, 7065–7077 (2001).
26. Takayanagi, H. *et al.* Induction and activation of the transcription factor NFATc1 (NFAT2) integrate RANKL signaling in terminal differentiation of osteoclasts. *Dev. Cell* **3**, 889–901 (2002).
27. Koga, T. *et al.* Costimulatory signals mediated by the ITAM motif cooperate with RANKL for bone homeostasis. *Nature* **428**, 758–763 (2004).
28. Takahashi, T. & Strittmatter, S. M. Plexin1 autoinhibition by the plexin sema domain. *Neuron* **29**, 429–439 (2001).
29. Narazaki, M. & Tosato, G. Ligand-induced internalization selects use of common receptor neuropilin-1 by VEGF165 and semaphorin3A. *Blood* **107**, 3892–3901 (2006).
30. Nishikawa, K. *et al.* Maf promotes osteoblast differentiation in mice by mediating the age-related switch in mesenchymal cell differentiation. *J. Clin. Invest.* **120**, 3455–3465 (2010).
31. Gimble, J. M., Zvonick, S., Floyd, Z. E., Kassem, M. & Nuttall, M. E. Playing with bone and fat. *J. Cell. Biochem.* **98**, 251–266 (2006).
32. Krishnan, V., Bryant, H. U. & Macdougald, O. A. Regulation of bone mass by Wnt signaling. *J. Clin. Invest.* **116**, 1202–1209 (2006).
33. Takada, I., Kouzmenko, A. P. & Kato, S. Wnt and PPAR γ signaling in osteoblastogenesis and adipogenesis. *Nature Rev. Rheumatol.* **5**, 442–447 (2009).
34. Toyofuku, T. *et al.* FARP2 triggers signals for Sema3A-mediated axonal repulsion. *Nature Neurosci.* **8**, 1712–1719 (2005).
35. Wu, X. *et al.* Rac1 activation controls nuclear localization of β -catenin during canonical Wnt signaling. *Cell* **133**, 340–353 (2008).
36. Takegahara, N. *et al.* Integral roles of a guanine nucleotide exchange factor, FARP2, in osteoclast podosome rearrangements. *FASEB J.* **24**, 4782–4792 (2010).
37. Nagashima, M. *et al.* Bisphosphonate (YM529) delays the repair of cortical bone defect after drill-hole injury by reducing terminal differentiation of osteoblasts in the mouse femur. *Bone* **36**, 502–511 (2005).
38. Tang, Y. *et al.* TGF- β 1-induced migration of bone mesenchymal stem cells couples bone resorption with formation. *Nature Med.* **15**, 757–765 (2009).
39. Hayden, J. M., Mohan, S. & Baylink, D. J. The insulin-like growth factor system and the coupling of formation to resorption. *Bone* **17**, S93–S98 (1995).
40. Kawai, M., Mödder, U. I., Khosla, S. & Rosen, C. J. Emerging therapeutic opportunities for skeletal restoration. *Nature Rev. Regen. Med.* **10**, 141–156 (2011).
41. Grigoriadis, A. E. *et al.* c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
42. Asagiri, M. *et al.* Autoamplification of NFATc1 expression determines its essential role in bone homeostasis. *J. Exp. Med.* **202**, 1261–1269 (2005).
43. Mizuno, A. *et al.* Severe osteoporosis in mice lacking osteoclastogenesis inhibitory factor/osteoprotegerin. *Biochem. Biophys. Res. Commun.* **247**, 610–615 (1998).
44. Kaifu, T. *et al.* Osteopetrosis and thalamic hypomyelination with synaptic degeneration in DAP12-deficient mice. *J. Clin. Invest.* **111**, 323–332 (2003).
45. Hayashi, M. *et al.* Ly49Q, an ITIM-bearing NK receptor, positively regulates osteoclast differentiation. *Biochem. Biophys. Res. Commun.* **393**, 432–438 (2010).
46. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. D. Ginty and A. L. Kolodkin for providing the *Nrp1*^{Sema-} knockin mice. We thank Y. Goshima for providing vectors and technical help. We thank A. Yamaguchi, H. Asahara and F. Suto for providing reagents and technical help. We also thank K. Okamoto, T. Negishi-Koga, K. Nishikawa, H. Inoue, T. Suda, T. Ando, Y. Kunisawa, Y. Ogihara and S. Fukuse for discussion and assistance. This work was supported in part by a grant for the Exploratory Research for Advanced Technology Program, the Takayanagi Osteonetwork Project from the Japan Science and Technology Agency; Grant-in-Aid for Young Scientist A from the Japan Society for the Promotion of Science (JSPS); a Grant-in-Aid for Challenging Exploratory Research from the JSPS; grants for the Global Center of Excellence Program from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and grants from the Tokyo Biochemical Research Foundation, the Life Science Foundation of Japan, Takeda Science Foundation, Uehara Memorial Foundation, Naito Foundation, BMKK RA Research Fund and Astellas Foundation for Research on Metabolic Disorders.

Author Contributions M.H. performed most of the experiments, interpreted the results and prepared the manuscript. T.N. performed immunohistochemical experiments and provided advice on project planning and data interpretation and prepared the manuscript. M.T. provided technical help. T.K. conducted the GeneChip analysis. A.K. provided advice on project planning and technical help. H.T. directed, supervised the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.T. (taka.csi@tmd.ac.jp).

METHODS

Cell culture. For *in vitro* osteoclast differentiation in the monoculture system, primary bone marrow cells (1×10^5 cells per cm^2) were suspended in culture medium (α -MEM containing penicillin, streptomycin and 10% FBS) supplemented with 10 ng ml^{-1} M-CSF (R&D Systems) for two days to obtain BMs. The resultant BMs were further cultured in medium supplemented with 10 ng ml^{-1} M-CSF and $5\text{--}50 \text{ ng ml}^{-1}$ RANKL (PeproTech) for three days. Culture medium was changed every second day. Where indicated, calvarial cell-conditioned medium with or without soluble Nr1p (R&D Systems) was added 12 h before or 12 h after the RANKL stimulation. Sema3A-Fc (R&D Systems) was added 12 h before or 12 h after the RANKL stimulation. For Sema6D treatment, BMs were collected and seeded onto culture plates coated with soluble recombinant Sema6D (R&D Systems), and cultured with 10 ng ml^{-1} M-CSF and 5 ng ml^{-1} RANKL for three days. For the generation of osteoclast *in vitro* in the coculture system, primary bone marrow cells (5×10^4 cells per cm^2) and calvarial cells (5×10^3 cells per cm^2) were cultured in the presence of 10 nM $1\alpha,25\text{-dihydroxyvitamin D}_3$ and $1 \mu\text{M}$ prostaglandin E2 for 4–6 days. For human osteoclast differentiation, human peripheral blood mononuclear cells were separated from peripheral blood obtained from healthy volunteers by density gradient centrifugation with Lymphoprep (AXIS-SHIELD). Cells (2×10^5 cells per 0.5 ml) were cultured in α -MEM with 10% FBS supplemented with 30 ng ml^{-1} M-CSF for two days. The resultant preosteoclasts were further cultured in medium supplemented with 30 ng ml^{-1} M-CSF and 60 ng ml^{-1} RANKL for four days. Culture medium was changed every second day. The differentiation of osteoclasts was evaluated by TRAP staining. A NF- κ B activation inhibitor (6-amino-4-(phenoxyphenylethylamino)quinazoline; Calbiochem) was used to inhibit NF- κ B activity. The concentration of intracellular calcium was measured as described²⁷. For *in vitro* osteoblast differentiation, calvarial cells were isolated from the calvarial bone of newborn mice by enzymatic digestion in α -MEM with 0.1% collagenase and 0.2% dispase, and were cultured with α -MEM with 10% FBS. After two days, cells were reseeded (1×10^4 cells per cm^2) and cultured with osteogenic medium (100 mM ascorbic acid, 5 mM β -glycerophosphate and 10 nM dexamethasone). Culture medium was changed every third day. After seven days, ALP staining and activity measurement were performed, and after 21 days, bone nodule formation was assessed by alizarin red staining. Human mesenchymal stem cells (Lonza) were cultured according to the manufacturer's protocol. To induce adipocyte differentiation *in vitro*, primary bone marrow cells (5×10^5 cells per cm^2) were cultured with α -MEM containing 10% FBS. After 24 h, non-adherent cells were removed and adherent cells were cultured with adipogenic medium (0.5 mM 3-isobutyl-1-methylxanthine, $5 \mu\text{g ml}^{-1}$ insulin and $1 \mu\text{M}$ dexamethasone) for 10–14 days. Culture medium was changed every second day. Lipid accumulation in adipocytes was determined with Oil Red O staining. Cell proliferation was determined using a cell proliferation ELISA kit (Roche Applied Science). The percentage of apoptotic cells was determined by TUNEL (TdT-mediated dUTP nick end labelling) staining with the MEBSTAIN apoptosis kit direct (MBL). For colony forming unit (CFU) assays, primary bone marrow cells (2.5×10^5 cells per cm^2) were seeded and cultured with MesenCult basal medium supplemented with mesenchymal stem cell stimulatory supplements (StemCell Technologies). On day 10, the cells were stained with toluidine blue. For CFU-ALP and CFU-osteoblast (CFU-OB) assays, primary bone marrow cells (2.5×10^5 cells per cm^2) were seeded and cultured with MesenCult basal medium and mesenchymal stem cell stimulatory supplements plus 100 mM ascorbic acid, 5 mM β -glycerophosphate and 10 nM dexamethasone. On day 10, CFU-ALP colonies were stained for ALP. Mineral deposition was determined with von Kossa staining of CFU-OB colonies on day 25. Primary osteoblasts and osteocytes were isolated from the long bones of CAG-CAT-EGFP/Dmp1-Cre double transgenic mice as described¹¹.

Purification and identification of inhibitory factor of osteoclast differentiation. Mouse calvarial cells were statically cultured with α -MEM supplemented with 1% FBS. Medium was conditioned for 72 h and filtered to remove non-adherent cells and debris. Conditioned medium was concentrated by ammonium sulphate precipitation (40% saturation), and the pellet was suspended in 10 mM sodium phosphate buffer, pH 7.4, and desalted using a PD-10 column (GE Healthcare). Concentrated conditioned media were then loaded onto a Mono Q 5/50 column (GE Healthcare) in 10 mM sodium phosphate buffer, pH 7.4, and proteins binding to the Mono Q matrix were eluted by a gradient of 0–100% 1 M NaCl and 10 mM sodium phosphate buffer, pH 7.4. Protein concentrations were determined by the absorbance at 280 nm. The effect of the fractionated conditioned media on osteoclast differentiation was examined by directly adding each fraction to RANKL-induced osteoclast differentiation *in vitro*. Fractions 33–35 contained a high concentration of NaCl, which exerts an inhibitory effect on osteoclast differentiation. Proteins of the highly inhibitory fractions were dissolved in SDS-PAGE sample buffer (Nacalai Tesque) and the sample was resolved by SDS-PAGE. Protein bands were visualized by Coomassie brilliant blue staining and all the protein bands were excised by scalpel. The samples were

analysed using nano-LC-MS/MS by Japan Bioservice. The data were submitted to the MASCOT program for identification.

Immunohistochemical staining. After fixation in 4% paraformaldehyde, bone tissues were decalcified in 10% EDTA at 4°C for 2 weeks and embedded in paraffin after dehydration. For immunohistochemical staining, antigen retrieval was carried out with 10 mM citric acid, pH 6.0, at room temperature for 2 h. After quenching of endogenous peroxidase activity by incubation with 3% H_2O_2 in methanol, the sections were incubated with an anti-Sema3A polyclonal antibody (Santa Cruz Biotechnology) in immunoreaction enhancer solution (Can Get Signal immunostain, TOYOBO) at 4°C for overnight. After washing with PBS, the sections were incubated with peroxidase-conjugated secondary antibody according to the manufacturer's instructions (histofine, Nichirei Bioscience). The signals were visualized with 3,3'-diaminobenzidine tetrahydrochloride and H_2O_2 . TRAP staining was conducted after the immunostaining. Haematoxylin was used for nuclear counterstaining.

Western blot and immunoprecipitation analyses. Cell lysate or culture supernatant of calvarial cells was subjected to western blot analysis using the specific antibodies for Nr1p (Calbiochem), β -actin (Sigma-Aldrich), Sema3A, Nr1p, p50, p65, histone H1 (Santa Cruz Biotechnology), phospho-ERK, ERK, phospho-JNK, JNK, phospho-p38, p38, phospho-IKK α/β , IKK α , IKK β , phospho-PLC γ 2, PLC γ 2 (Cell Signaling Technology), β -catenin (Millipore) and α -tubulin (MBL). Nuclear proteins were prepared with nuclear extract kit in accordance with the manufacturer's protocol (Active Motif). For immunoprecipitation analysis, cells were solubilized in lysis buffer (1% Nonidet P-40 in 50 mM NaCl, 50 mM Tris-HCl, 5 mM EDTA, 1 mM NaF and 2 mM PMSF), supplemented with complete protease inhibitor cocktail (Roche Applied Science). Immunoprecipitation was performed by incubation with an anti-Flag M2 (Sigma-Aldrich) or anti-PlxnA1 antibody (Santa Cruz Biotechnology) followed by the addition of dynabeads protein G (Invitrogen). Immune complexes were separated by electrophoresis followed by blotting with anti-Flag M2, anti-V5 (Invitrogen), anti-Nr1p, anti-PlxnA1, anti-DAP12 (Santa Cruz Biotechnology) and anti-TREM2 antibodies (R&D Systems).

Flow cytometric analysis. For the analysis of bone marrow-derived osteoclast precursor cells, a single cell suspension of mouse bone marrow cells was stained with anti-CD3 ϵ (145-2C11, eBioscience), anti-B220 (RA3-6B2, eBioscience), anti-CD11b (M1/70, eBioscience), anti-CD115 (AFS98, eBioscience) and anti-CD117 (2B8, eBioscience) antibodies. For intracellular staining of BMs, anti-CD11b, anti-CD115 and anti-Nr1p (R&D Systems) were used. Flow cytometric analysis was performed using FACSCantoII with Diva software (BD Biosciences).

ELISA. Concentrations of soluble RANKL and Opg in serum were determined using ELISA kits (R&D Systems), according to the manufacturer's instruction.

Retroviral gene transfer. The retroviral vector pMXs-Nr1p-IRES-EGFP was constructed by inserting DNA fragments encoding Nr1p into pMXs-IRES-EGFP. The construction of the retroviral vectors pMX-FARP2-IRES-GFP and pMX- Δ GEF-FARP2-IRES-GFP was described previously³⁶. For the construction of the retroviral vectors pSIREN-RetroQ-ZsGreen-shNr1p and pSIREN-RetroQ-ZsGreen-shControl, RNA targeting regions with a hairpin sequence (Nr1p shRNA sense: 5'-GCCCGAATGTTCTCAGAACTACGAGTAGTCTGA GAACATTCGGGCTTTT-3'; Nr1p shRNA antisense: 5'-AAAAAGCCCGA ATGTTCTCAGAACTACGAGTAGTCTGAGAACATTCGGGC-3'; control shRNA sense: 5'-GTGCGTTGCTAGTACCAACTTCAAGAGATTTTACGC GT-3'; control shRNA antisense: 5'-ACGCGTAAATAATCTCTTGAGTTGG TACTAGCAACGCAC-3') were inserted into RNAi-ready pSIREN-RetroQ-ZsGreen (Clontech). The retrovirus supernatants were obtained by transfecting the retroviral vectors into the Plat-E packaging cell line using FuGENE 6 (Roche Applied Science).

Chromatin immunoprecipitation assay. Chromatin immunoprecipitation assay was performed using the ChIP-IT express chromatin immunoprecipitation kit (Active Motif) according to the manufacturer's instructions. The antibodies used for immunoprecipitation were anti-p50, anti-p65 and normal rabbit IgG (Santa Cruz Biotechnology). The primer sequences were as follows: Nr1p region 1, 5'-CATACGTGACCTTGGCTCT-3' and 5'-CCTGGCTGGAGATTCAGAGA-3'; Nr1p region 2, 5'-ACCTTACCCACCAGCTCCTT-3' and 5'-ATACG CCACCACTTACGAG-3'; Nr1p region 3, 5'-ATGTGGCTTGGTGAAAG GAG-3' 5'-TGCTTCTACCTTCGGGTGAT-3'.

Reporter gene assay. The reporter plasmid Nr1p-Luc was constructed by sub-cloning a 3,259 base pair fragment of the 5' flanking region of the mouse Nr1p gene into the pGL3-basic vector (Promega). The reporter plasmids and the expression plasmids were transfected into NIH3T3 cells using FuGENE 6 (Roche Applied Science). After 36 h, dual luciferase assay was performed according to the manufacturer's protocol (Promega).

Migration assay. BMs suspended in complete medium were added to the upper chamber of transwell units (Corning). Inserts were placed into the lower chambers

of transwell units containing M-CSF with or without Sema3A-Fc. After incubation, cells were fixed with 4% paraformaldehyde and stained with 0.5% toluidine blue. The cells on the upper side of the membrane were removed and the cells that had migrated to the lower side of the membrane and chamber were counted.

G-LISA small G protein activation assay. RhoA and Rac GTPase activation were determined using the G-LISA RhoA and Rac absorbance-based assay (Cytoskeleton) according to the manufacturer's instructions. In brief, cell lysates were prepared and normalized. After the addition of antibodies against RhoA or Rac and the incubation with the horseradish peroxidase detection reagent, signals were detected with a spectrophotometer.

In vivo treatment with recombinant Sema3A. Five-week-old C57BL/6 mice were given weekly intravenous injections of 1 mg per kg body weight of Sema3A-Fc or vehicle for four weeks. Three days after the last injection, bone analysis was performed as described earlier.

Bone regeneration model. Skeletal injury was generated as described previously³⁷. In brief, C57BL/6 mice were anaesthetized with an intraperitoneal

injection of pentobarbital sodium. A 5-mm longitudinal incision was made over the proximal femur and the bone surface was exposed by splitting the muscle. A 0.5-mm hole was made by drilling through the anterior portion of the diaphysis of the bilateral femurs. After four and seven days of surgery, femoral defects were treated with Sema3A-Fc (0.5 mg per kg body weight) by injection into the injury site. Mice were euthanized at day 14 after surgery and bone analyses were performed.

Ovariectomy-induced bone loss. Nine-week-old female mice were ovariectomized or sham operated. More than five mice were examined in each group. Ovariectomized mice were given weekly intravenous injections of 1 mg per kg body weight of Sema3A-Fc or vehicle for four weeks. Three days after the last injection, all of the mice were euthanized and subjected to bone analysis as described earlier.

Statistical analyses. Statistical analyses were performed using the unpaired two-tailed Student's *t* test (**P* < 0.05; ***P* < 0.01; ****P* < 0.001; NS, not significant, throughout the paper). All data are expressed as the mean ± s.e.m. Results are representative examples of more than three independent experiments.

A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants

Elke Barbez^{1,2}, Martin Kubeš³, Jakub Rolčík⁴, Chloé Béziat^{1,2}, Aleš Pěňčík⁵, Bangjun Wang⁶, Michel Ruiz Rosquete^{1,2}, Jinsheng Zhu⁶, Petre I. Dobrev³, Yuree Lee⁷, Eva Zajímalová³, Jan Petrášek³, Markus Geisler⁶, Jiří Friml¹ & Jürgen Kleine-Vehn^{1,2}

The phytohormone auxin acts as a prominent signal, providing, by its local accumulation or depletion in selected cells, a spatial and temporal reference for changes in the developmental program^{1–7}. The distribution of auxin depends on both auxin metabolism (biosynthesis, conjugation and degradation)^{8–10} and cellular auxin transport^{11–15}. We identified *in silico* a novel putative auxin transport facilitator family, called PIN-LIKES (PILS). Here we illustrate that PILS proteins are required for auxin-dependent regulation of plant growth by determining the cellular sensitivity to auxin. PILS proteins regulate intracellular auxin accumulation at the endoplasmic reticulum and thus auxin availability for nuclear auxin signalling. PILS activity affects the level of endogenous auxin indole-3-acetic acid (IAA), presumably via intracellular accumulation and metabolism. Our findings reveal that the transport machinery to compartmentalize auxin within the cell is of an unexpected molecular complexity and demonstrate this compartmentalization to be functionally important for a number of developmental processes.

Prominent auxin carriers with fundamental importance during plant development are PIN-FORMED (PIN) proteins^{1–3,6,9,15}. PIN1-type auxin carriers regulate the directional intercellular auxin transport at the plasma membrane. In contrast, atypical family member PIN5 regulates intracellular auxin compartmentalization into the lumen of the endoplasmic reticulum and its role in auxin homeostasis was recently identified^{15,16}. PIN proteins have a predicted central hydrophilic loop, flanked at each side by five transmembrane domains. We screened *in silico* for novel PIN-like putative carrier proteins with a predicted topology similar to PIN proteins (Fig. 1a and Supplementary Fig. 2) and identified a protein family of seven members (Fig. 1b) in *Arabidopsis thaliana*, which we designated as the PILS proteins. In contrast to the similarities in the predicted protein topology, PIN and PILS proteins do not show pronounced protein sequence identity (10–18%), which limits the identification of PILS proteins by conventional, reciprocal basic local alignment search tool (BLAST) approaches. However, the distinct PIN and PILS protein families contain both the InterPro auxin carrier domain which is an *in-silico*-defined domain, aiming to predict auxin transport function (<http://www.ebi.ac.uk/panda/InterPro.html>). The PILS putative carrier family is conserved throughout the whole plant lineage, including unicellular algae (such as *Ostreococcus tauri* and *Chlamydomonas reinhardtii*) (Supplementary Fig. 3) where PIN proteins are absent¹⁶, indicating that PILS proteins are evolutionarily older.

PILS genes are broadly expressed in various tissues (Fig. 1c) and *PILS2–PILS7* were transcriptionally upregulated by auxin application in wild-type seedlings (Fig. 1d–f and Supplementary Fig. 4), indicating a role in auxin-dependent processes. To investigate the potential function of the putative PILS auxin flux facilitators in plant development, we overexpressed PILS proteins using the constitutive, viral 35S promoter. Ectopic expression of PILS genes, such as *PILS1* or *PILS3*,

resulted in dwarfed and/or bushy plants showing severe defects in flower development, leading to sterility in the T1 generation (Fig. 2a–d). Flowers of these PILS-overexpressing plants showed severe patterning defects, such as homeotic transformation of flower organs into new flower buds, triplication of the gynoecium or unfused carpels (Fig. 2b–d). To circumvent sterility, we screened for weaker *p35S::PILS* lines and isolated moderately *PILS5* overexpressing lines showing fertile flower development.

To assess further the developmental importance and potential redundancy of PILS proteins in auxin-regulated processes, we focused

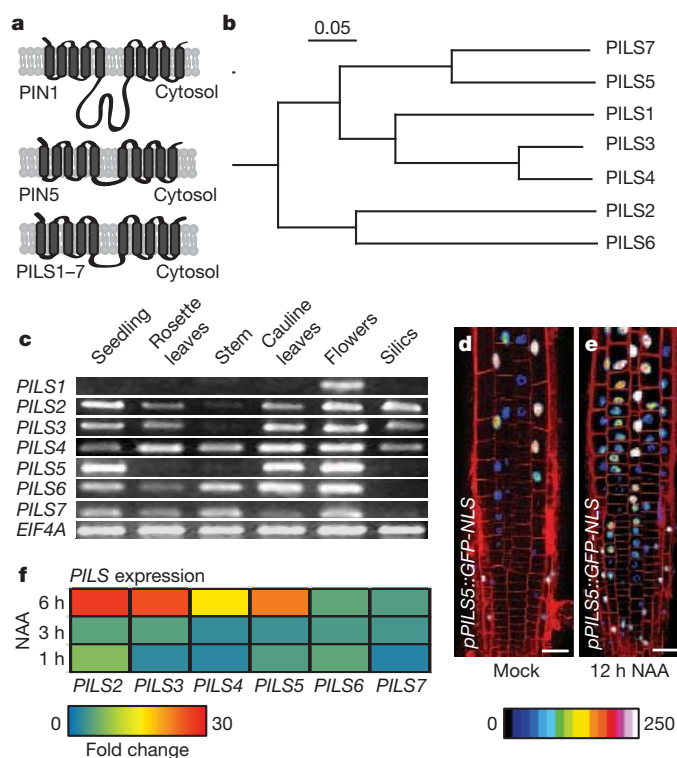
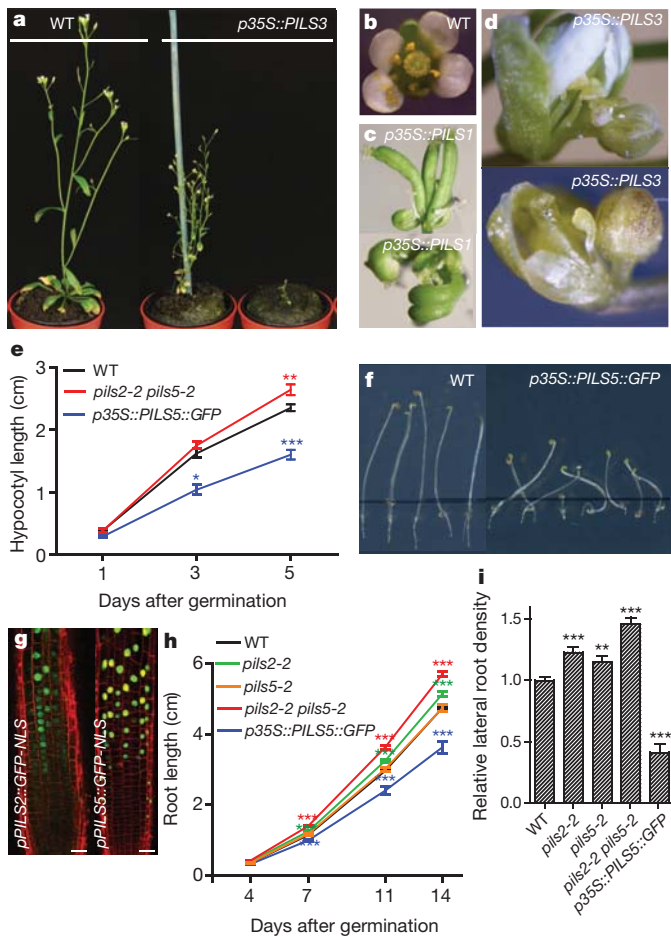


Figure 1 | Novel PILS protein family. **a**, Predicted topology of both PIN1-/PIN5-type PIN subfamilies and PILS proteins. **b**, Phylogenetic tree of the *A. thaliana* PILS proteins. Scale depicts 0.05 substitutions per position. **c**, Reverse transcription PCR (RT-PCR) of the seven PILS genes and *EIF4A* (control) in several plant tissues. **d**, **e**, 1-Naphthaleneacetic acid (NAA; 10 μM, 12 h)-induced *pPILS5::GFP-NLS* (GFP fused to nuclear localisation signal, NLS) expression in the root transition zone. Colour-code (black to white) depicts (low to high) GFP signal intensity. Propidium-iodide-stained cell walls in red. Scale bar, 25 μm. **f**, Quantitative RT-PCR of *PILS2–7* after 10 μM NAA treatment for 1, 3 and 6 h. Colour code from blue (low) to red (high) depicts fold changes (0–30) (see also Supplementary Fig. 4).

¹Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University, 9002 Ghent, Belgium. ²Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences (BOKU), 1190 Vienna, Austria. ³Institute of Experimental Botany, The Academy of Sciences of the Czech Republic, 16502 Praha 6, Czech Republic. ⁴Laboratory of Growth Regulators, Faculty of Science, Palacký University and Institute of Experimental Botany AS CR, Šlechtitělská 11, 78371 Olomouc, Czech Republic. ⁵Centre of the Region Haná for Biotechnological and Agricultural Research, Department of Growth Regulators, Faculty of Science, Palacký University, Šlechtitělská 11, 78371 Olomouc, Czech Republic. ⁶Department of Biology - Plant Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland. ⁷Department of Plant Molecular Biology, University of Lausanne, Quartier Sorge, 1015 Lausanne, Switzerland.

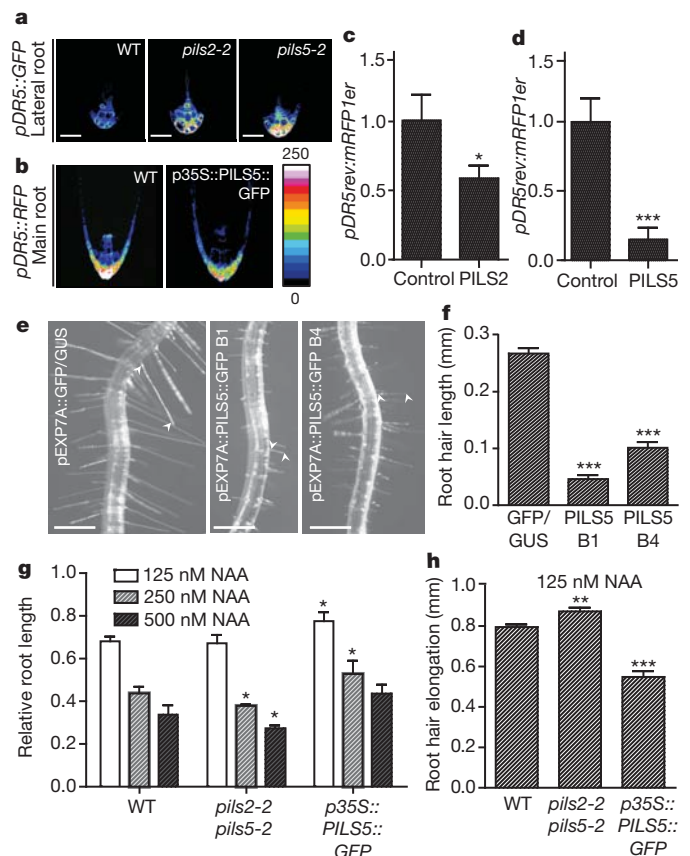


on *PILS2* and *PILS5*, because they are the most abundantly expressed *PILS* genes in seedlings (Fig. 1c) and display partially overlapping expression domains (Fig. 2g). Initially, we investigated *PILS* function in auxin-dependent hypocotyl and root growth. Dark-grown *p35S::PILS5::GFP*-expressing (coding for a *PILS5*–green fluorescent protein (GFP) fusion) and *pils2 pils5* double-mutant seedlings showed reduced and enhanced hypocotyl growth, respectively (Fig. 2e and Supplementary Fig. 5). *PILS5* gain-of-function also resulted in agravitropic hypocotyl growth (Fig. 2f). *PILS2* and *PILS5* showed a particular overlapping expression in the root transition zone (Fig. 2g), suggesting a redundant role in regulating root growth. Indeed, *pils2* single-mutant and more pronounced *pils2 pils5* double-mutant seedlings showed significantly longer roots compared to wild-type seedlings, whereas seedlings overexpressing *PILS5* had shorter roots (Fig. 2h). Collectively, our data indicate the requirement of defined *PILS* protein activity for auxin-dependent growth regulation.

Beside the root and shoot organ growth, auxin tightly controls *de novo* organ formation such as lateral root organogenesis¹. Intriguingly, *pils2* and *pils5* single-mutants and more pronounced *pils2 pils5* double-mutants showed higher lateral root density (Fig. 2i). On the contrary, *PILS5* gain-of-function reduced lateral rooting (Fig. 2i).

These findings indicate developmental importance of *PILS* proteins in auxin-regulated processes, such as *de novo* organ formation and growth regulation.

Next we investigated whether auxin responses are affected in *pils* mutants using the auxin response reporter DR5 (ref. 17). *pils2-2* knockdown and *pils5-2* knockout mutants did not show altered DR5 activity in the main root tips (Supplementary Fig. 6a), but showed higher *pDR5rev::GFP* signal intensity in lateral roots (Fig. 3a). Moderately *p35S::PILS5::GFP*-expressing seedlings showed a visibly reduced auxin response maximum in the very root tip of main (Fig. 3b) and lateral roots (Supplementary Fig. 6b). To distinguish between direct and indirect effects, we investigated the effect of *PILS* proteins on nuclear auxin signalling at the cellular level. Therefore, we transiently co-expressed *PILS2* or *PILS5* together with the auxin response reporter *pDR5rev::mRFP1er* (expressing the red fluorescent protein (RFP) in response to auxin signalling) in tobacco Bright Yellow 2 (BY-2) cells. *PILS2* or *PILS5* expression in BY-2 cells reduced auxin signalling as visualized by *pDR5rev::mRFP1er* activity (Fig. 3c, d). These findings indicate that *PILS* proteins negatively affect nuclear auxin signalling, presumably by affecting cellular auxin homeostasis.



To address whether cellular PILS action on auxin signalling affects cellular growth, we expressed *PILS1*, *PILS3* and *PILS5* under a root-hair-specific promoter. Deviations in free (active) indole-3-acetic acid (IAA) levels or in auxin signalling induce (high levels) or repress (low levels) root hair growth¹⁸. As expected, the root-hair-specific expression of *PILS1*, *PILS3* and *PILS5* significantly reduced root hair length (Fig. 3e, f and Supplementary Fig. 7a–c), possibly due to PILS-dependent regulation of auxin homeostasis and signalling.

Next we tested whether PILS proteins affect auxin-dependent cellular growth responses. We treated PILS gain- and loss-of-function mutants with auxin that inhibits primary root growth. *pils2 pils5* loss- and *PILS5* gain-of-function mutants showed hyper- and hypersensitive root growth, respectively (Fig. 3g). In agreement with these observations, the auxin-promoted root hair growth was also enhanced in *pils2 pils5* double-mutants and reduced in *PILS5* overexpressors (Fig. 3h). This set of data indicates that PILS putative auxin facilitators modulate auxin-dependent growth responses during plant development.

To unravel the mechanism by which PILS proteins regulate auxin-dependent plant development, we investigated the subcellular localization of PILS proteins. We introduced amino- or carboxy-terminal GREEN/RED FLUORESCENT PROTEIN (GFP/RFP) fusions with PILS proteins and transiently or stably expressed these fusion proteins in tobacco BY-2 cell culture, *Arabidopsis* seedlings and heterologously in yeast (Fig. 4a–c and Supplementary Fig. 8). *PILS1*, *PILS2*, *PILS3*, *PILS5*, *PILS6* and *PILS7* GFP/RFP fusion proteins localized to the endoplasmic reticulum in all analysed systems (Fig. 4a–c and Supplementary Figs 8 and 10c). In contrast, N- and C-terminal *PILS4* fusions did not show detectable fluorescence. The transgenic *pPILS5::PILS5::GFP* also showed endoplasmic reticulum localization (Fig. 4a) and complemented the *pils2 pils5* double-mutant to the *pils2* single-mutant level (Supplementary Fig. 9), indicating PILS function at the endoplasmic reticulum.

Next, we addressed whether putative PILS auxin carriers affect cellular auxin accumulation. We generated *PILS2* oestradiol-inducible tobacco BY-2 cell cultures and performed ³H-IAA accumulation assays. *PILS2* induction increased the accumulation of radioactivity in BY-2 cells (Fig. 4d). In accordance with the auxin accumulation assays in BY-2 cells, *pils2 pils5* double-mutant protoplasts showed significantly higher auxin export (Fig. 4e and Supplementary Fig. 10a), indicating reduced auxin retention capacity in *pils2 pils5* loss-of-function mutants. These gain and loss-of-function studies consistently illustrate that *PILS2* and *PILS5* function at the endoplasmic reticulum controls cellular accumulation of auxin.

To additionally address PILS protein function in a non-plant system, we expressed *PILS3*, *PILS5* and *PILS7* (tagged to GFP or haemagglutinin (HA)) in *Saccharomyces cerevisiae* yeast cells. In accordance with the effect of *PILS2* and *PILS5* in plant cells, the expression of *A. thaliana PILS3*, *PILS5* and *PILS7* in yeast also increased retention of exogenously applied auxin (Fig. 4f and Supplementary Fig. 10b). To assess the specificity to auxin we used the common organic control benzoic acid (BA). Notably, *PILS3*, *PILS5* and *PILS7* did not affect the cellular accumulation of BA (Fig. 4f and Supplementary Fig. 10b). These findings indicate that PILS proteins specifically regulate cellular auxin accumulation.

PILS proteins increase cellular auxin accumulation, but decrease auxin signalling, which is reminiscent of the regulation of auxin metabolism shown for the PIN5 auxin transporter¹⁶. To assess the potential action of PILS proteins on auxin metabolism, we analysed the effect of PILS proteins on the auxin metabolism using the inducible *PILS2* BY-2 cell line. The high-performance liquid chromatography (HPLC) chromatogram of IAA showed a *PILS2*-dependent reduction of free IAA within 20 min (Fig. 4g, h), indicating that *PILS2* activity affects auxin metabolism.

In accordance with the auxin metabolite profiling in BY-2 cell cultures, *pils2*, *pils5* and *pils2 pils5* mutant *Arabidopsis* seedlings showed significantly higher free IAA levels compared to wild-type seedlings (Fig. 4i). Intriguingly, the ratio of amid auxin conjugates IAA-Glutamate (Glu)

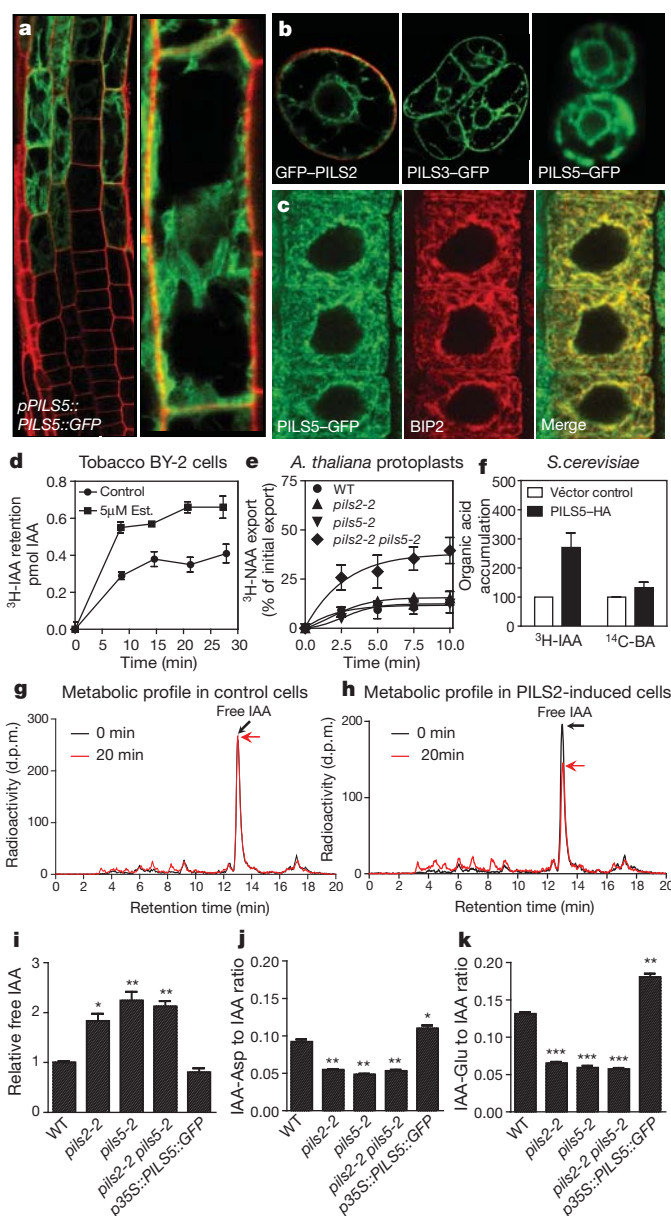


Figure 4 | PILS involvement in cellular auxin homeostasis. a–c, GFP–*PILS2* (b), *PILS3*–GFP (b), *PILS5*–GFP (a–c) under endogenous (a) or constitutive promoter (b, c) in *A. thaliana* root (a, c), tobacco BY-2 cells (b, left panels) and *Saccharomyces cerevisiae* yeast cells (b, right panel). c, Immunocytochemistry of *PILS5*–GFP and the endoplasmic reticulum marker BIP2. d, ³H-IAA retention in tobacco BY-2 cells upon oestradiol (Est.)-induced *PILS2* expression (*n* = 3 repetitions). e, ³H-NAA export assay in *A. thaliana* protoplasts of wild type (WT), *pils2-2*, *pils5-2* and *pils2 pils5* mutants (*n* = 3 repetitions; see also Supplementary Fig. 10a). f, Accumulation of auxin and benzoic acid in *Saccharomyces cerevisiae* yeast cells transformed with *pGPD::PILS5::HA* or an empty vector (*n* = 3 repetitions; see also Supplementary Fig. 10b). g, h, ³H-IAA metabolic profile (HPLC chromatogram) in *PILS2* non-oestradiol-induced (g) and oestradiol-induced (h) tobacco BY-2 cells after 0 (black line) and 20 min (red line). i–k, Liquid chromatography-mass spectrometry (LC-MS)-derived free IAA levels (i) and ratios of IAA-Asp (j) and IAA-Glu (k) to free IAA in WT, *pils2-2*, *pils5-2*, *pils2 pils5* and *p35S::PILS5::GFP* (*n* = 3 repetitions). Error bars represent s.e.m. Student *t*-test *P*-values: **P* < 0.05, ***P* < 0.001, ****P* < 0.0001. Scale bar, 10 μm.

and IAA-Aspartate (Asp) to free IAA was also significantly shifted towards free IAA in *pils2* and *pils5* loss-of-function mutants (Fig. 4j, k). Moderately *p35S::PILS5::GFP*-expressing seedlings showed a visibly reduced auxin response maximum in the very root tip (Fig. 3b), but only a mild and statistically non significant reduction in free IAA levels

at the whole seedling level (Fig. 4i). However, the ratio of IAA-Glu and IAA-Asp to free IAA was significantly shifted towards the conjugates in *p35S::PILS5::GFP* seedlings (Fig. 4j, k), indicating a higher rate of auxin conjugation.

These findings are indicative of a PILS function in cellular auxin homeostasis by regulating auxin metabolism. Hence, we propose a model in which PILS proteins at the endoplasmic reticulum membrane facilitate intracellular auxin accumulation, which seems to contribute to the possibly compartmentalized regulation of auxin metabolism (Supplementary Fig. 1). It is tempting to speculate that auxin conjugation could also take place in the endoplasmic reticulum, although the molecular components remain to be identified.

Our *in silico* and reverse genetics approaches led to the identification of a novel family of putative auxin transport facilitators. All our genetic, pharmacological, cell biological, physiological and biochemical approaches consistently suggest that PILS proteins function at the endoplasmic reticulum membrane, regulate intracellular auxin accumulation and affect free IAA levels, presumably through conjugation-based auxin metabolism. The PILS action on cellular auxin homeostasis is reminiscent of the function of the atypical PIN family member PIN5 (ref. 16). We uncovered that an additional, distinct protein family regulates intracellular auxin homeostasis. Our analyses of the PILS proteins suggest that intracellular auxin transport and, hence, auxin compartmentalization might be evolutionarily older than directional, cell-to-cell PIN-dependent auxin transport mechanisms. The identification of a novel protein family for the regulation of intracellular auxin homeostasis highlights the evolutionary and developmental importance of intracellular auxin transport. Further studies will address the potential interplay or possible diversified function of endoplasmic reticulum localized PIN5 and PILS1–7 proteins.

METHODS SUMMARY

Plant material and growth conditions. We used *Arabidopsis thaliana* of ecotype Columbia 0 (Col-0). The *Nicotiana tabacum* L. cv. Bright Yellow 2 (BY-2) cell line¹⁹ was used as suspension-cultured cells. *pils2-1* (SALK_024808), *pils2-2* (SALK_125391), *pils5-1* (SALK_070653) and *pils5-2* (SALK_072996) were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). Gateway cloning was used to construct *pPILS2::GFP/GUS*, *pPILS5::GFP/GUS*, *p35S::PILS1-7*, *p35S::GFP::PILS1-7*, *p35S::PILS1-7::GFP*, *pPILS5::PILS5::GFP* and *pMDC7_B(pUBQ)::PILS2*. The following lines and constructs have been previously described: *pDR5rev::mRFP1er²⁰*, *pDR5rev::GFP²* and *p35S::GFP::HDEL²¹*. Seeds were stratified at 4 °C for 2 days in the dark. Seedlings were grown vertically on half Murashige and Skoog medium. Plants were grown under long-day (16 h light/8 h dark) conditions at 20–22 °C.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 December 2011; accepted 24 February 2012.

Published online 15 April 2012.

1. Benková, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003).
2. Friml, J. *et al.* Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*. *Nature* **426**, 147–153 (2003).
3. Reinhardt, D. *et al.* Regulation of phyllotaxis by polar auxin transport. *Nature* **426**, 255–260 (2003).

4. Leyser, O. Dynamic integration of auxin transport and signalling. *Curr. Biol.* **16**, R424–R433 (2006).
5. Dubrovsky, J. G. *et al.* Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc. Natl Acad. Sci. USA* **105**, 8790–8794 (2008).
6. Sorefan, K. *et al.* A regulated auxin minimum is required for seed dispersal in *Arabidopsis*. *Nature* **459**, 583–586 (2009).
7. Prasad, K. *et al.* *Arabidopsis* PLETHORA transcription factors control phyllotaxis. *Curr. Biol.* **21**, 1123–1128 (2011).
8. Woodward, A. W. & Bartel, B. Auxin: regulation, action, and interaction. *Ann. Bot.* **95**, 707–735 (2005).
9. Ikeda, Y. *et al.* Local auxin biosynthesis modulates gradient-directed planar polarity in *Arabidopsis*. *Nature Cell Biol.* **11**, 731–738 (2009).
10. Zhao, Y. Auxin biosynthesis and its role in plant development. *Annu. Rev. Plant Biol.* **61**, 49–64 (2010).
11. Bennett, M. J. *et al.* *Arabidopsis* AUX1 gene: a permease-like regulator of root gravitropism. *Science* **273**, 948–950 (1996).
12. Luschig, C., Gaxiola, R. A., Grisafi, P. & Fink, G. R. EIR1, a root-specific protein involved in auxin transport, is required for gravitropism in *Arabidopsis thaliana*. *Genes Dev.* **12**, 2175–2187 (1998).
13. Geisler, M. *et al.* Cellular efflux of auxin catalyzed by the *Arabidopsis* MDR/PGP transporter AtPGP1. *Plant J.* **44**, 179–194 (2005).
14. Petrášek, J. *et al.* PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science* **312**, 914–918 (2006).
15. Zajímalová, E., Murphy, A. S., Yang, H., Hoyerová, K. & Hošek, P. Auxin transporters—why so many? *Cold Spring Harb. Perspect. Biol.* **2**, a001552 (2010).
16. Mravec, J. *et al.* ER-localized PIN5 auxin transporter mediates subcellular homeostasis of phytohormone auxin. *Nature* **459**, 1136–1140 (2009).
17. Ulmasov, T., Murfett, J., Hagen, G. & Guilfoyle, T. J. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* **9**, 1963–1971 (1997).
18. Lee, S. H. & Cho, H. T. PINOID positively regulates auxin efflux in *Arabidopsis* root hair cells and tobacco cells. *Plant Cell* **18**, 1604–1616 (2006).
19. Nagata, T., Nemoto, Y. & Hasegawa, S. Tobacco BY-2 cell line as the “HeLa” cells in the cell biology of higher plants. *Int. Rev. Cytol.* **132**, 1–30 (1992).
20. Marin, E. *et al.* miR390, *Arabidopsis* TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell* **22**, 1104–1117 (2010).
21. Langhans, M. *et al.* *In vivo* trafficking and localization of p24 proteins in plant cells. *Traffic* **9**, 770–785 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to C. Braeckman for plant transformation; W. Ardiles for sequencing support; L. Charrier for technical assistance; A. Maizel, N. Geldner and P. Pimpl for providing material; J.K.-V. group members for critical reading of the manuscript and the BOKU-VIBT Imaging Center for access and expertise. This work was supported by the Vienna Science and Technology Fund (WWTF) (to J.K.-V.), the Agency for Innovation by Science and Technology (IWT) (predoctoral fellowship to E.B.), the Odysseus program of the Research Foundation-Flanders (to J.F.), the Swiss National Funds (to M.G.), the Ministry of Education, Youth and Sports of the Czech Republic (LC06034) (to E.Z.), Grant Agency of the Czech Republic project P305/11/2476 (to J.P.) and P305/11/0797 (to E.Z.).

Author Contributions E.B. and J.K.V. conceived the project. E.B. carried out most of the experiments. M.K., P.I.B., E.Z. and J.P. performed auxin metabolite profile and auxin accumulation in BY-2. C.B. analysed auxin-dependent PILS expression and contributed to phenotype analysis. M.R.R. contributed to PILS cloning. J.R. and A.P. measured auxin content in *Arabidopsis*. B.W., J.Z. and M.G. performed auxin accumulation in yeast and protoplasts. Y.L. modified the oestradiol-inducible vector. E.B., M.K., J.R., E.Z., J.P., M.G., J.F. and J.K.V. discussed the experimental procedures. All authors analysed and discussed the data; E.B. and J.K.V. wrote the paper and all authors saw and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.K.-V. (juergen.kleine-vehn@boku.ac.at).

METHODS

PILS gene accession codes. Sequence data from this article can be found in The *Arabidopsis* Information Resource (TAIR; <http://www.arabidopsis.org/>) or GenBank/EMBL databases under the following accession numbers: *PILS1* (At1g20925), *PILS2* (At1g71090), *PILS3* (At1g76520), *PILS4* (At1g76530), *PILS5* (At2g17500), *PILS6* (At5g01990), *PILS7* (At5g65980).

Plant material, growth conditions and DNA constructs. We used *Arabidopsis thaliana* of ecotype Columbia 0 (Col-0). The *Nicotiana tabacum* L. cv. Bright Yellow 2 (BY-2) cell line¹⁹ was used as suspension-cultured cells. *pils2-1* (SALK_024808), *pils2-2* (SALK_125391), *pils5-1* (SALK_070653) and *pils5-2* (SALK_072996) were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). Insertion sites were verified, homozygous lines selected and the decrease or absence of the respective *PILS* transcript was shown by RT-PCR. The *pils2-2* and the *pils5-2* mutants were crossed into *DR5rev::GFP*². Gateway cloning was used to construct *pPILS2::GFP::GUS*, *pPILS5::GFP::GUS*, *p35S::PILS1-7*, *p35S::GFP::PILS1-7*, *p35S::PILS1-7::GFP*, *pPILS5::PILS5::GFP* and *pMDC7_B(pUBQ)::PILS2*. The *PILS* full-length genomic fragments, complementary DNA and promoter regions were amplified by PCR from genomic DNA and cDNA, respectively. The PCR was performed using the high fidelity DNA polymerase "I proof" (Bio-Rad). The primers used are given below. The full genomic and cDNA fragments were cloned into the pDONR221 (Invitrogen) vector and the promoter regions into pDONR-P4P1 using Invitrogen BP-clonase according to manufacturer's instructions. Coding sequences were transferred from the entry clones to gateway-compatible destination vectors (given below) using the Invitrogen LR clonase(+) according to manufacturer's instructions. The resulting constructs were transformed into Col-0 plants by floral dipping in *Agrobacterium tumefaciens* liquid cultures. Yeast vectors were transformed into budding yeast (*Saccharomyces cerevisiae*) via electroporation. The *p35S::PILS5::GFP* line was crossed into *pDR5rev::mRFP1er*²⁰. The following lines and constructs have been previously described: *pDR5rev::mRFP1er*²⁰, *pDR5rev::GFP*² and *p35S::GFP::HDEL*²¹. Seeds were stratified at 4 °C for 2 days in the dark. Seedlings were grown vertically on half Murashige and Skoog medium. Plants were grown under long-day (16 h light/8 h dark) conditions at 20–22 °C.

Chemicals. 1-Naphthaleneacetic acid (NAA) was supplied by Duchefa, 2,4-dichlorophenoxy acetic acid, oestradiol and propidium iodide by Sigma-Aldrich and ³H-indole-3-acetic acid (³H-IAA), ³H-naphthalene-1-acetic acid (³H-NAA) and ¹⁴C-benzoic acid (¹⁴C-BA) (specific radioactivity 20 Ci mmol⁻¹) by American Radiolabelled Chemicals.

RNA extraction and quantitative real time PCR (qPCR). Whole RNA of seedlings was extracted using the RNeasy Mini Kit (Qiagen) in technical triplicates, the extracted RNA samples were treated with DNase (Ambion). qPCR analysis was performed using ICycler (Bio-Rad) with the Platinum SYBR Green qPCR Super-UDG kit (Invitrogen) following recommendations of the manufacturer. qPCR was carried out in 96-well optical reaction plates heated for 10 min to 95 °C to activate hot-start *Taq* DNA polymerase, followed by 40 cycles of denaturation for 60 s at 95 °C and annealing-extension for 60 s at 58 °C. Target quantifications were performed with specific primer pairs (given later) designed using Beacon Designer 4.0 (Premier Biosoft International). Expression levels were normalized to the expression levels of translation initiation factor EIF4A. The primers used are given later.

Phenotype analysis. For analysis of the root length and lateral root density, plates were scanned on a flat-bed scanner. Root hairs were imaged with a binocular microscope (Leica). For hypocotyls analysis, seeds on plates were exposed to light for 3 h at 18 °C, and cultivated in the dark at 20 °C. Seedlings were imaged in real time with an infrared camera (Canon) to define the exact moment of germination and analysed 1, 3 and 5 days after germination. Hypocotyls, root and root hair lengths were measured with the ImageJ (<http://rsb.info.nih.gov/ij/>) software. Lateral root density for each seedling was obtained by calculating the number of lateral roots per root length unit 14 days after germination. For analysis of hypocotyls length, a minimum of 15 hypocotyls per condition or mutant line were analysed in each experiment. For analysis of root length and lateral root density, a minimum of 40 plants per condition or mutant line were analysed in each experiment. Means and standard errors were calculated and the statistical significance was evaluated by the student *t*-test using the GraphPad Prism5 (<http://www.graphpad.com>) software. For the analysis of root hair growth, 20 seedlings per transgenic line were imaged by binocular (Leica) and 20 root hairs (randomly chosen in the root hair zone) per seedling were measured with the ImageJ (<http://rsb.info.nih.gov/ij/>) software. The mean and standard error of the mean per transgenic line were calculated and the statistical significance was evaluated by the student *t*-test. To obtain the auxin-dependent root hair elongation, the same number of root hairs per seedling, seedlings per mutant line and condition were analysed as described above. The untreated mean average root hair length of the respective genotype was subtracted from the individual auxin-treated root hair length to obtain auxin induced growth in millimetre. The mean and standard error of the mean of the respective genotype were calculated and the

statistical significance was evaluated by the student *t*-test using the GraphPad Prism5 (<http://www.graphpad.com>) software. All experiments were performed in at least three independent biological repetitions.

BY-2 plant material. Cells of tobacco line BY-2 (*Nicotiana tabacum* L., cv. Bright Yellow 2)¹⁹ transformed with *pMDC7_B(pUBQ)::PILS2* were cultured in liquid cultivation medium (3% (w/v) sucrose, 4.3 g l⁻¹ Murashige and Skoog salts, 100 mg l⁻¹ inositol, 1 mg l⁻¹ thiamine, 0.2 mg l⁻¹ 2,4-dichlorophenoxy acetic acid, and 200 mg l⁻¹ KH₂PO₄ (pH 5.8)). BY-2 cell lines were cultivated in darkness at 26 °C on an orbital incubator (Sanyo Gallenkamp, Schöeller Instruments; 150 r.p.m., 32-mm orbit) and subcultured weekly. Stock BY-2 calli were maintained on media solidified with 0.6% (w/v) agar and subcultured monthly.

Transient transformation of BY-2 cells and monitoring of cellular auxin signalling in BY2. Ten ml of three-day-old cells were harvested on filter paper by vacuum filtration and kept on plates with solid BY-2 medium. The cells were transformed via particle bombardment with a PDS 1000/He biolistic system (Bio-Rad) according to the manufacturer's instructions (http://www.bio-rad.com/webroot/web/pdf/lsl/literature/Bulletin_9075.pdf). 2 µl of plasmid DNA (0.05 µg µl⁻¹ of the *pDR5rev::mRFP1er* construct and 0.1 µg µl⁻¹ of *p35S::PILS2* and *p35S::PILS5*) was added to 6.25 µl of 1.6-µm diameter gold particles (dissolved in 50% glycerol). The suspension was supplemented with 2.5 µl spermidine (0.1 M stock solution) and 6.25 µl CaCl₂ (2.5 M stock solution). The particles were pelleted by centrifugation, washed twice with 70% and 100% ethanol and, subsequently, resuspended in 10 µl of 100% ethanol. Cells were bombarded under a pressure of 1,100 pounds per square inch. The plates were sealed with Parafilm and kept in the dark for 18 h at 25 °C. For microscopic analysis, cells were gently transferred from the filter to a microscopy slide (in water) and subsequently covered with a cover slip. Samples were analysed via confocal microscopy. The *pDR5rev::mRFP1er* expression was evaluated by defining the mean grey value (MGV) of each imaged cell (middle sections). For each experiment, confocal settings were defined based on the *pDR5rev::mRFP1er* signal of the control cells and remained unchanged during the respective experiments. Transformants were identified on the basis of the fluorescence of both proteins and imaged with a ×40 objective. Every experiment/transformation was done in triplicate and for each condition a total number of at least 60 transformed cells were imaged. For each experiment, the means and standard errors were calculated and the statistical significance (independence between the two populations) was obtained by the student *t*-test using the GraphPad Prism5 (<http://www.graphpad.com>) software.

Immunocytochemistry. Whole-mount immunological staining on 5-day-old seedlings was done in an Intavis robot according to the described protocol²². The antibodies used at the final dilutions were monoclonal mouse anti-BIP2 (Hsc70) at 1:200 (Stressgen Bioreagents), monoclonal rabbit anti-GFP at 1:600 (Invitrogen). The secondary anti-mouse (Invitrogen) and anti-rabbit (Sigma-Aldrich) antibodies conjugated with Cy3 and Alexa488 respectively were used at 1:600 dilution.

Microscopy. Confocal microscopy was done with a Zeiss 710 microscope (Zeiss) or Leica SP5 (Leica). Fluorescence signals for GFP (excitation 488 nm, emission peak 509 nm), mRFP1 (excitation 561 nm, emission peak 607 nm) and propidium iodide (PI) staining (excitation 536 nm, emission peak 617 nm) were detected with a ×20, ×40 (water immersion) or ×63 (water immersion) objective. Sequential scanning was used for double labelling to avoid crosstalk between channels. Fluorescence signal intensity was analysed with ImageJ (<http://rsb.info.nih.gov/ij/>) software and data were statistically evaluated with Microsoft Excel 2007.

Auxin transport assays in tobacco BY-2 cells, baker's yeast and *Arabidopsis thaliana* protoplasts. Auxin accumulation with 2-day-old BY-2 cells was measured as previously described^{14,23}. The ³H-IAA was added to give a final concentration of 2 nM. Accumulation results were expressed as pmols of particular auxin accumulated per 10⁶ cells. The 0.5-ml aliquots of cell suspension were collected continuously and accumulation of label was terminated by rapid filtration under reduced pressure on 22-mm-diameter cellulose filters. The cell cakes and filters were transferred to scintillation vials, extracted in 0.5 ml of 96% ethanol for 30 min, and afterwards 4 ml of scintillation solution (EcoLite Liquid Scintillation Fluid, MP Biomedicals) was added. Radioactivity was determined by liquid scintillation counter Packard Tri-Carb 2900TR (Packard-Canberra, Meridian). Yeast ³H-IAA loading was quantified with the unspecific ¹⁴C-benzoic acid as control assayed in parallel and performed as previously described²⁴. Relative export is calculated from yeast-retained radioactivity as follows: ((radioactivity in the yeast at time *t* = 10 min) – (radioactivity in the yeast at time *t* = 0)) × (100%)/(radioactivity in the yeast at *t* = 0 min). Unspecific loading due to diffusion was eliminated by vector control subtraction. IAA export from *Arabidopsis thaliana* mesophyll protoplasts was analysed as described¹⁶.

HPLC metabolic profiling in tobacco BY-2 cells. Two-days-old BY-2 cells were prepared for the experiment by equilibration in uptake buffer as already described for accumulation assays¹⁴. Experiments were done in uptake buffer and under

standard cultivation conditions. Cells were incubated with addition of 20 nM ³H-IAA for a period of 0 and 20 min. Cells and media (uptake buffer) were collected and frozen in liquid nitrogen (100 mg of fresh weight and 5 ml per sample). Extraction and purification of auxin metabolites in cells and media were performed as described^{25,26}. The metabolites were separated on HPLC consisting of autosampler and 235C diode array detector (Perkin Elmer), column Luna C18 (2), 150 × 4.6 mm, 3 μm (Phenomenex, Torrance, USA), mobile phase A: 40 mM CH₃COONH₄ (pH 4.0) and mobile phase B: CH₃CN/CH₃OH, 1/1, (v/v). Flow rate was 0.6 ml min⁻¹ with linear gradient 30–50% B for 10 min, 50–100% B for 1 min, 100% B for 2 min, 10–30% B for 1 min. The column eluate was monitored on 235C DAD followed by Ramona 2000 flow-through radioactivity detector (Raytest GmbH) after online mixing with three volumes (1.8 ml min⁻¹) of liquid scintillation cocktail (Flo-Scint III, Perkin Elmer). The radioactive metabolites were identified on the basis of comparison of their retention times with authentic standards. For the results presentation the total integrated area of chromatogram plots has been normalized based on the equalization of total accumulated radiolabel.

In silico and phylogenetic analysis. PILS genes were identified via the SMART-protein tool from EMBL (<http://smart.embl-heidelberg.de/>)^{27,28}. Phylogenetic tree of AtPILS was constructed with the DNA-man software version 4.0. PILS topologies were defined by the online HMM-top tool (<http://www.enzim.hu/hmmtop/>)²⁹ and visualized by the TMRPres2D software (<http://biophysics.biol.uoa.gr/TMRPres2D/download.jsp>)³⁰. PILS orthologues were identified with the online tool Plaza (<http://bioinformatics.psb.ugent.be/plaza/>)³¹.

Free IAA and conjugate measurements in *Arabidopsis thaliana*. For the quantification of free IAA and its amino acid conjugates, approximately 10 mg of plant material was taken into analysis. The samples were processed as previously described³² and quantified by UHPLC-MS/MS.

Used primers and vectors. Genotyping primers: pils2-1 RP, CTGGAGAAACC TGACATCTCG; pils2-1 LP, GATTGAAGCCGGCTTAAATTC; pils2-2 RP, CTGGAGAAACCTGACATCTCG; pils2-2 LP, TACCATTGATCTGTCTTCG GG; pils5-1 RP, TTGAGACCCGTATCATTGGAG; pils5-1 LP, TGTCTG ATAAACCTTTTCAGG; pils5-2 RP, TACTGCACCGAAAATGAAACC; pils5-2 LP, TTGTACTATTTCACCGGCTC.

Insert primer (Lbb) (combine with RP): GCGTGGACCGCTTGCTGCAACT. RT-PCR primers used for insertion lines: PILS2 Fw, GCGATCATTATCGGATC AGT; PILS2 Rev, TTGCATACCTTGACAGTAGTC; PILS5 Fw, TGTGTA AGCCCGTAATCCATGAAC; PILS5 Rev, TTCATTGCGGACCCTTTAAT CAGC.

qPCR primers: PILS1 Fw, CGGTAAACACAGCTCCACTTC; PILS1 Rev, GCAACAAGTAACGCACAACC; PILS2 Fw, GTGATGCTTGTACTTGGTGG TATG; PILS2 Rev, AACTTGAACATTGGATCTGCTGAG; PILS3 Fw, AGGCG ACCATGCAAGTGTG; PILS3 Rev, GTGGTACAGCTAGATGACAGTGAG; PILS4 Fw, TGTCATACTAAGCCTCCTTCAC; PILS4 Rev, CTCGCAACTCTC AGAATCTCC; PILS5 Fw, CTTGGAATAGTCTGTGTTTCGGTAC; PILS5 Rev, GCACTGAGCATTCGCTTTCAG; PILS6 Fw, GCCTACATCAGTGCTCTCAG; PILS6 Rev, GCACTGAGCATTCGCTTTCAG; PILS7 Fw, TCCTCCAGACCCTC TCTTTCG; PILS7 Rev, ACAAGAAGATGACCGAGCACTC; Eif4a Fw, CTGGAGGTTTTCAGGCTGGTAT; Eif4a Rev, CCAAGGTTGAAAGCAAG AAGA.

Cloning primers: gDNA/cDNA: PILS1_Fw, GGGGACAAGTTTGTACAAAA AAGCAGGCTCGATGAGGATGAGGCTTTTGGATC; PILS1_Rev, GGGGAC CACTTTGTACAAGAAAGCTGGGTC(TCA)GGCTACGAGCCACATGAAG AATG; PILS2_Fw, GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGT CAGGTTTCTCCAGTGGAA; PILS2_Rev, GGGGACCACTTTGTACAAGAAA GCTGGGTC(TCA)TTGCATACCTTGACAGTAGTCTC; PILS3_Fw, GGGG ACAAGTTTGTACAAAAAAGCAGGCTCGATGGTGAAGCTTTTGGAGCTG

TTC; PILS3_Rev, GGGGACCACTTTGTACAAGAAAGCTGGGTC(TCA)AGC TACAAGCCACATGAAGAAATG; PILS4_Fw, GGGGACAAGTTTGTACAAAA AAGCAGGCTCGATGAAGCTTTTGGAGTTGTTC; PILS4_Rev, GGGGAC CACTTTGTACAAGAAAGCTGGGTC(TCA)TGTCACAAGCCACATGAAGA ATG; PILS5_Fw, GGGGACAAGTTTGTACAAAAAAGCAGGCTCGATGGG ATTCTGGTCTGTTGTGGA; PILS5_Rev, GGGGACCACTTTGTACAAGAAA GCTGGGTC(TCA)GACTAACAAGTGAAGGAAGATGG; PILS6_Fw, GGGGA CAAGTTTGTACAAAAAAGCAGGCTCGATGATTGCTCGGATCCTTGCCG; PILS6_Rev, GGGGACCACTTTGTACAAGAAAGCTGGGTC(TCA)GAAGAG TATGTTAATGTAGAGTAC; PILS7_Fw, GGGGACAAGTTTGTACAAAAA GCAGGCTCGATGGGTTTCTTAGAGTTGTTGGA; PILS7_Rev, GGGGACCA CTTTGTACAAGAAAGCTGGGTC(TCA)GGAGAGGATGGAGAGGAAGAT GG.

Promoter: PILS2_Fw, GGGGACAAGTTTGTATAGAAAAAGTTGCGAACTCC ATTGTTAACAGTAATAGC; PILS2_Rev, GGGGACTGCTTTTTTGTACAAA CTTGCCTCGATCTCACTATGTAAAGCTCG; PILS5_Fw, GGGGACAAGTT TGTATAGAAAAAGTTGCGCAATATACGTGACGTGGTCCACT; PILS5_Rev, GGGGACTGCTTTTTTGTACAAACTTGCCTTTTTATGTGGTTCTTTAGAC. Destination vectors: *pPILS::GFP/GUS*: *pKGWFS7*, *p35S::PILS::GFP*: *pH7FWG2.0*, *p35S::GFP::PILS*: *pH7WGF2.0*, *p35S::PILS::RFP*: *pK7RWG2.0*, *p35S::RFP::PILS*: *pK7WGR2.0* and *p35S::PILS_D*: *pH7WG2D.1* (ref. 33), and *pPILS::PILS::GFP*: *pK7m34GW.0* (ref. 34).

Oestradiol-inducible PILS: *pMDC7_B(pUBQ)*³⁵ (*p35S* promoter was exchanged by the *pUBQ10* promoter).

22. Sauer, M., Paciorek, T., Benkova, E. & Friml, J. Immunocytochemical techniques for whole-mount *in situ* protein localization in plants. *Nature Protocols* **1**, 98–103 (2006).
23. Delbarre, A., Muller, P., Imhoff, V. & Guern, J. Comparison of mechanisms controlling uptake and accumulation of 2,4-dichlorophenoxy acetic acid, naphthalene-1-acetic acid, and indole-3-acetic acid in suspension-cultured tobacco cells. *Planta* **198**, 532–541 (1996).
24. Bailly, A. *et al.* Modulation of P-glycoproteins by auxin transport inhibitors is mediated by interaction with immunophilins. *J. Biol. Chem.* **283**, 21817–21826 (2008).
25. Dobrev, P. I., Havlíček, L., Vágner, M., Malbeck, J. & Kamínek, M. Purification and determination of plant hormones auxin and abscisic acid using solid phase extraction and two-dimensional high performance liquid chromatography. *J. Chromatogr. A* **1075**, 159–166 (2005).
26. Dobrev, P. I. & Kamínek, M. Fast and efficient separation of cytokinins from auxin and abscisic acid and their purification using mixed-mode solid-phase extraction. *J. Chromatogr. A* **950**, 21–29 (2002).
27. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
28. Letunic, I., Doerks, T. & Bork, P. SMART 6: recent updates and new developments. *Nucleic Acids Res.* **37**, 229–232 (2009).
29. Tusnády, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).
30. Spyropoulos, I. C., Liakopoulos, T. D., Bagos, P. G. & Hamodrakas, S. J. TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics* **20**, 3258–3260 (2004).
31. Proost, S. *et al.* PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**, 3718–3731 (2009).
32. Pěnčík, A. *et al.* Isolation of novel indole-3-acetic acid conjugates by immunoaffinity extraction. *Talanta* **80**, 651–655 (2009).
33. Karimi, M., De Meyer, B. & Hilson, P. Modular cloning in plant cells. *Trends Plant Sci.* **10**, 103–105 (2005).
34. Karimi, M., Inze, D. & Depicker, A. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195 (2002).
35. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in plants. *Plant Physiol.* **133**, 462–469 (2003).

Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia

Catherine C. Smith¹, Qi Wang², Chen-Shan Chin³, Sara Salerno¹, Lauren E. Damon¹, Mark J. Levis⁴, Alexander E. Perl⁵, Kevin J. Travers³, Susana Wang³, Jeremy P. Hunt⁶†, Patrick P. Zarrinkar⁶†, Eric E. Schadt³†, Andrew Kasarskis³†, John Kuriyan² & Neil P. Shah^{1,7}

Effective targeted cancer therapeutic development depends upon distinguishing disease-associated ‘driver’ mutations, which have causative roles in malignancy pathogenesis, from ‘passenger’ mutations, which are dispensable for cancer initiation and maintenance. Translational studies of clinically active targeted therapeutics can definitively discriminate driver from passenger lesions and provide valuable insights into human cancer biology. Activating internal tandem duplication (ITD) mutations in *FLT3* (*FLT3-ITD*) are detected in approximately 20% of acute myeloid leukaemia (AML) patients and are associated with a poor prognosis¹. Abundant scientific² and clinical evidence^{1,3}, including the lack of convincing clinical activity of early *FLT3* inhibitors^{4,5}, suggests that *FLT3-ITD* probably represents a passenger lesion. Here we report point mutations at three residues within the kinase domain of *FLT3-ITD* that confer substantial *in vitro* resistance to AC220 (quizartinib), an active investigational inhibitor of *FLT3*, *KIT*, *PDGFRA*, *PDGFRB* and *RET*^{6,7}; evolution of AC220-resistant substitutions at two of these amino acid positions was observed in eight of eight *FLT3-ITD*-positive AML patients with acquired resistance to AC220. Our findings demonstrate that *FLT3-ITD* can represent a driver lesion and valid therapeutic target in human AML. AC220-resistant *FLT3* kinase domain mutants represent high-value targets for future *FLT3* inhibitor development efforts.

Perhaps the most compelling evidence so far that *FLT3-ITD* could represent a driver mutation in AML was the identification of a secondary *FLT3* kinase domain mutation that conferred moderate resistance to the multikinase inhibitor PKC412 in a single *FLT3-ITD*⁺ patient who relapsed after an initial bone marrow response⁸. Although the broad-spectrum kinase inhibitor sorafenib has anecdotally achieved bone marrow remissions in *FLT3-ITD*⁺ AML patients⁹, whether its mechanism of action is mediated through inhibition of *FLT3* or a distinct kinase is unclear. Indeed, two patients who relapsed after an initial response to sorafenib had no detectable *FLT3* kinase domain mutations at the time of resistance¹⁰.

A recent interim analysis of 53 relapsed/refractory *FLT3-ITD*⁺ AML patients evaluable for efficacy in a multinational phase II trial of AC220 monotherapy documented a composite complete remission (<5% bone marrow blasts) rate of 45% (frequently associated with incomplete recovery of peripheral blood counts)⁷. We sought to use the clinical activity of AC220 as a tool to define *FLT3-ITD* as a driver or passenger mutation in human AML. Using a previously validated *in vitro* saturation mutagenesis assay¹¹, we identified AC220 resistance-conferring mutations at four residues in the kinase domain of *FLT3-ITD* (Fig. 1a). Mutations at three of these amino acid positions conferred high degrees of *in vitro* AC220 resistance as demonstrated in proliferation (Fig. 1b) and cell-based biochemical assays (Fig. 1c). These residues consist of the ‘gatekeeper’ residue (F691) and two

residues within the activation loop (D835, Y842). For unclear reasons, the E608K substitution did not confer substantial AC220 resistance and was not further characterized.

We next assessed the presence of drug-resistant kinase domain mutations in *FLT3-ITD* in eight paired pre-treatment and relapse samples obtained from *FLT3-ITD*⁺ AML patients who initially achieved morphological reduction of bone marrow blasts to ≤5% with AC220 monotherapy, but subsequently relapsed despite continued AC220 treatment. In every case, subcloning and sequencing¹² of individual *FLT3-ITD* alleles revealed mutations at the time of relapse (Table 1) that were not detected pre-treatment (Supplementary Table 1). Mutations were confined to two of the three critical residues identified in our *in vitro* screen. The activation loop mutation D835Y was detected in three cases, D835V in two, and the gatekeeper mutation F691L was identified in three. Additionally, one novel activation loop mutation, D835F, was identified in a single patient. This mutation confers substantial *in vitro* resistance to AC220 (Supplementary Fig. 1) and cross-resistance to sorafenib (data not shown), and was probably

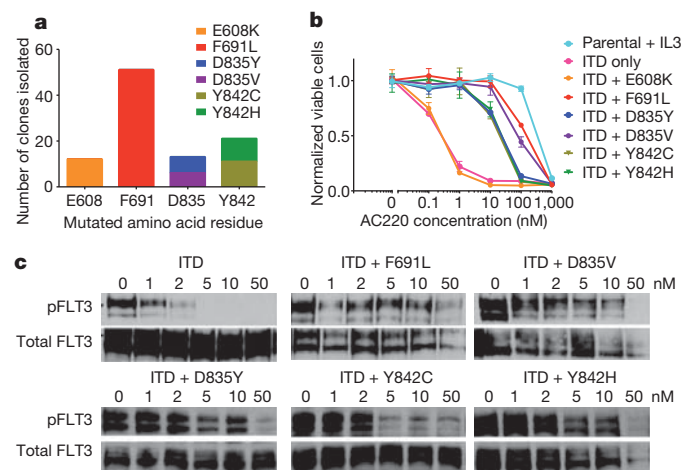


Figure 1 | Mutation screen of *FLT3-ITD* reveals secondary kinase domain mutations that cause varying degrees of resistance to AC220. **a**, Numbers of independent AC220-resistant Ba/F3 *FLT3-ITD* subpopulations with amino acid substitution at the indicated residue obtained from a saturation mutagenesis assay ($n = 97$ clones). **b**, Normalized cell viability of Ba/F3 populations stably expressing *FLT3-ITD* mutant isoforms after 48 h in various concentrations of AC220 (error bars represent standard deviations of triplicates from the same experiment). **c**, Western blot analysis using anti-phospho-*FLT3* (pFLT3) or anti-*FLT3* antibody performed on lysates from IL-3-independent Ba/F3 populations expressing the *FLT3-ITD* mutant isoforms indicated. Cells were exposed to AC220 at the indicated concentrations for 90 min.

¹Division of Hematology/Oncology, University of California, San Francisco, California 94143, USA. ²Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ³Pacific Biosciences, Menlo Park, California 94025, USA. ⁴Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland 21231, USA. ⁵Abramson Cancer Center of the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁶Ambit Biosciences, San Diego, California 92121, USA. ⁷Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94115, USA. [†]Present addresses: KINOMEScan Division of DiscoverRx Corporation, San Diego, California 92121, USA (J.P.H.); Blueprint Medicines Corporation, Cambridge, Massachusetts 02142, USA (P.P.Z.); Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, New York 10029, USA (E.E.S. and A.K.).

Table 1 | Summary of FLT3 kinase domain mutations in patients relapsed on AC220

Subject number	Sex	Age (years)	Prior therapy	Karyotype at enrolment	Karyotype at relapse	Blasts in relapse sample (%)	New mutation at relapse	ITD ⁺ clones with mutation	Weeks on study
1009-003	F	75	7+3	45~54,XX,+3,+6,+7,+8,+13,+14,+21,+22[cp15]/46,XX[5]	52,XX,+3,+6,+7,+8,+10,+12,+13[cp7]/46,XX[14]	90	D835F	6/15	12
1011-006	M	70	7+3, low-dose cytarabine	Normal	ND	10	D835Y	4/15	8
1011-007	F	56	7+3, HAM	Normal	46,XX,del(11)(p?13p?15)[12]/46,XX[9]	80	F691L D835V	4/24 5/24	11
1005-004	F	60	Cytarabine and mitoxantrone	Normal	Normal	92	F691L	9/22	19
1005-006	M	43	7+3, MEC, allogeneic stem cell transplant	6,XY,t(1;15)(p22;q15)	ND	59	D835Y	8/17	6
1005-007	F	59	7+3, HDAC	Normal	ND	39	D835V	9/21	23
1005-009	M	68	Cytarabine and mitoxantrone	Normal	ND	58	D835Y	8/14	19
1005-010	M	52	7+3, HDAC, mitoxantrone and etoposide	46,XY,t(4;12)(q26;p11.2), t(8;14)(q13;q11.2)	ND	22	F691L	6/18	20

All patients achieved morphological bone marrow blasts of $\leq 5\%$ at best response. 7+3, low-dose cytarabine for 7 days plus 3 days anthracycline; HAM, high-dose cytarabine plus mitoxantrone; HDAC, high-dose cytarabine; MEC, mitoxantrone, etoposide, cytarabine. ND, not done.

not detected in our saturation mutagenesis screen because its creation requires a two-nucleotide substitution. One patient (1011-007) seemed to have evolved polyclonal resistance, with both F691L and D835V mutations detected on separate *FLT3-ITD* sequences. Collectively, these findings suggest that clinical response and relapse in each of these eight patients is mechanistically mediated through modulation of FLT3-ITD kinase activity.

To assess more precisely for resistance-conferring mutations at relapse, we used a recently described single molecule real-time (SMRT; Pacific Biosciences) sequencing platform, which can provide sequencing reads of sufficient length to enable focused interrogation of *FLT3-ITD* alleles (Supplementary Fig. 2)¹³. With this assay, hundreds of reads (range 19–930) spanning the ITD region and kinase domain with an average read length of greater than 1 kilobase (kb) (Supplementary Table 2) were reliably obtained from individual patient samples. Attention was focused on the amino acid codons identified in the *in vitro* screen for AC220 resistance-conferring mutations. SMRT sequencing confirmed the presence of resistance-conferring kinase domain mutations in *FLT3-ITD* at relapse in all eight patient samples (Table 2). Consistent with the results obtained by subcloning and sequencing, mutations at E608 and Y842 were not detected. The frequency of individual alternative codon substitutions within *FLT3-ITD* ranged from as low as 2.7% (D835F in patient 1005-006) to 50.6% (D835Y in patient 1005-009). The presence of polyclonal resistance was confirmed in patient 1011-007, and noted in an additional three cases: 1009-003, 1005-006 and 1005-007 (Table 2 and Supplementary Fig. 3). In general, mutations were detected on distinct

molecules, although in the case of 1011-007, a subset of *FLT3-ITD* molecules with F691L also harboured a D835V mutation (5/21 observations; 23.8% of *FLT3-ITD*(F691L) alleles; data not shown). Analysis of three normal control samples revealed base substitutions at these residues at a very low frequency (Table 2 and Supplementary Table 3). The evolution of polyclonal resistance due to secondary kinase domain mutations in *FLT3-ITD* in four of eight relapsed patients is supportive of a central dependence upon FLT3-ITD signalling in the leukaemic clone of a subset of AML patients, and indicative of profound selective pressure exerted upon this clone by AC220. Additionally, these findings reveal the genetic complexity of drug-resistant disease that may evolve in cancer patients on clinically active therapy.

All mutations identified at relapse were detected at a frequency significantly higher than that observed in a normal control, and although relapse occurred relatively rapidly in some patients, mutations were not convincingly detectable before treatment. The aggregate frequency of all mutations at relapse in individual patients ranged from approximately 20–50% in all cases, which is consistent with leukaemic blasts homozygous for *FLT3-ITD* and containing one drug-resistant allele per cell, although the presence of a heterogeneous blast population with only a subset of drug-resistant *FLT3-ITD*⁺ cells expressing kinase domain mutations cannot be excluded.

The five substitutions that conferred a high degree of resistance to AC220 *in vitro* were cross-resistant to sorafenib in cell-based growth and biochemical assays (Supplementary Fig. 4). The degree of relative resistance to sorafenib associated with these mutants was generally similar to that observed with AC220 (Supplementary Table 4).

Table 2 | Third-generation sequencing identifies polyclonal FLT3 kinase domain mutations

Subject number	Mutation	Native codon	Alternative codon	Pre-treatment		Relapse		Normal control no. 1	
				Observed alternative codon frequency in ITD ⁺ sequences (%)	Total number of ITD ⁺ sequences sampled	Observed alternative codon frequency in ITD ⁺ sequences (%)	Total number of ITD ⁺ sequences sampled	Observed alternative codon frequency (%)	Total number of sequences sampled
1009-003	D835Y	GAT	TAT	0.21	482	8.4	332	0.00	768
	D835V	GAT	GTT	0.00	482	3.3	332	0.13	768
	D835F	GAT	TTT	0.00	482	10.2	332	0.00	768
1011-006	D835Y	GAT	TAT	0.00	196	41.0	402	0.00	768
1011-007	F691L	TTT	TTG	0.18	561	6.2	341	0.22	450
	D835Y	GAT	TAT	0.00	930	3.0	436	0.00	768
	D835V	GAT	GTT	0.43	930	29.6	436	0.13	768
1005-004	F691L	TTT	TTG	0.00	496	29.6	513	0.22	450
1005-006	D835Y	GAT	TAT	0.00	171	39.5	261	0.00	768
	D835F	GAT	TTT	0.00	171	2.7	261	0.00	768
	D835Y	GAT	TAT	0.00	57	4.0	378	0.00	768
1005-007	D835Y	GAT	TAT	0.00	57	47.4	378	0.13	768
	D835V	GAT	GTT	0.00	57	50.6	445	0.00	768
	D835Y	GAT	TAT	0.00	19	25.3	150	0.22	450
1005-010	F691L	TTT	TTG	0.00	387				

All P values $< 1 \times 10^{-5}$ for alternative codon frequencies at relapse compared to a representative normal control sample (no. 1 refers to one of three normal control samples analysed).

To understand the structural effects of AC220-resistance conferring mutations, we modelled the binding of AC220 to FLT3 (Fig. 2a). The crystal structure of the FLT3 kinase domain has been determined previously in an inactive conformation¹⁴ that resembles the inactive conformations of ABL¹⁵, KIT¹⁶ and insulin receptor tyrosine kinase¹⁷, with the activation loop folded back onto the ATP-binding cleft (loop-in conformation), thereby preventing substrate loading. The Asp-Phe-Gly (DFG) motif at the amino-terminal end of the activation loop adopts the DFG-out conformation, in which the Asp side chain, which normally coordinates a magnesium ion, is removed from the active site. The activation of FLT3 would require flipping of the DFG motif and reorganization of the activation loop, as observed in ABL¹⁸ and insulin receptor kinase¹⁹. Previously published binding data suggest that AC220 is a type II kinase inhibitor that preferentially binds to the inactive, DFG-out kinase conformation²⁰. Our molecular docking analysis supports a model whereby AC220 interacts favourably with the DFG-out, inactive conformation. In the docked AC220–FLT3 model, the AC220 amide group is 2.6 Å from the carbonyl group of C694, consistent with the formation of a hydrogen bond. The phenol ring of AC220 forms a perpendicular aromatic–aromatic interaction²¹ with F830 in the DFG motif (Fig. 2b). This interaction would not be possible in the DFG-in, active kinase conformation. The gatekeeper residue F691 forms a π – π stacking contact with the benzo-imidazol-thiazol moiety of AC220, which may further stabilize the complex. Substitutions at F691 with non-aromatic residues such as leucine may not compensate for the π – π stacking interaction.

Residues D835 and Y842 stabilize the inactive conformation of the activation loop by forming hydrogen bonds with the main-chain amide group of S838 and the side chain of D811, respectively (Fig. 2c). Thus, replacement of either residue might destabilize this particular inactive conformation of the activation loop, which would then be expected to

hinder the binding of AC220. In further support of this model, a binding study of AC220 and sorafenib, a crystallographically proven type II inhibitor²², revealed that the binding affinity of both inhibitors to the FLT3 D835V mutant is substantially reduced compared to native FLT3, both in the presence and absence of the juxtamembrane domain containing the ITD (Supplementary Fig. 5 and Supplementary Table 5).

Other potential explanations for the mechanism of resistance conferred by AC220-resistant mutants include increased kinase activity and differential activation of downstream effectors. Western blot analysis of cells expressing AC220-resistant FLT3-ITD-mutant isoforms revealed increased FLT3 autophosphorylation of D835 mutant isoforms, but no discernable difference in phosphorylation of downstream targets, including the direct FLT3-ITD target STAT5A/B²³, and no difference in cellular proliferation (Supplementary Fig. 6). Overall, these data support a primarily structural mechanism for AC220 resistance mediated by mutations at F691, D835 and Y842, although further studies are necessary for definitive confirmation. We speculate that the ability to retain inhibitory activity against activation loop substitutions at D835 and Y842 will require a type I FLT3 kinase inhibitor capable of effectively binding to the active, DFG-in conformation of the kinase.

Substitutions at gatekeeper residues such as FLT3-ITD(F691) are well-documented causes of resistance to kinase inhibitors^{12,24}. Analogues of the FLT3-ITD(D835V) activation loop mutation have proven problematic for a number of kinase inhibitors: KIT(D816V), an activating mutation that is commonly detected in systemic mastocytosis, confers a high degree of resistance to imatinib and other KIT inhibitors²⁵. Our data, although derived from a small cohort of patients that will need to be validated in larger studies, suggest that substitutions at F691 and D835 in FLT3-ITD will pose substantial barriers to disease control in AML patients treated with either AC220 or sorafenib, and therefore represent high-value targets for novel FLT3 inhibitor development efforts.

Compelling data suggest that activating *FLT3* mutations are acquired relatively late during leukaemogenesis in a pre-established clone^{1,3}, and alone are insufficient to cause acute leukaemia in pre-clinical models². Recent evidence suggests that the molecular heterogeneity of individual leukaemias can be substantial, and can occur in both branching and linear fashions early during leukaemogenesis, including at the leukaemia-initiating or 'leukaemic stem' cell level²⁶. Collectively, our data are consistent with acquisition of *FLT3-ITD* and drug-resistant *FLT3* kinase domain mutations in a leukaemia-initiating cell population, although formal transplantation studies in mice are needed to address this question definitively. Our findings validate FLT3-ITD as a therapeutic target in human AML, and suggest that FLT3-ITD is capable of conferring a state of 'oncogene addiction', whereby cellular survival pathways associated with normal or precancerous cells can become hijacked, leading to a state of reliance upon key signalling molecules that can be exploited therapeutically. This work supports the exploration of therapeutic strategies targeting select activating mutations in other signalling molecules that are believed to be acquired relatively late in disease evolution, such as JAK2 (ref. 27) or RAS³, with agents capable of achieving clinically meaningful target inhibition. Further studies will be required to identify mechanisms of drug resistance that may circumvent reliance on activated FLT3 by activation of downstream or parallel pathways, as has been described with other kinase inhibitors²⁸. To that end, translational studies using detailed molecular analyses of primary samples obtained from AML patients treated with clinically effective targeted therapeutics promise to further inform mechanisms of drug resistance, strategies for future drug development, and models of disease evolution.

METHODS SUMMARY

MSCVpuroFLT3-ITD plasmid DNA was mutagenized and used to generate AC220-resistant Ba/F3 clones as previously described¹¹. The *FLT3* kinase domain

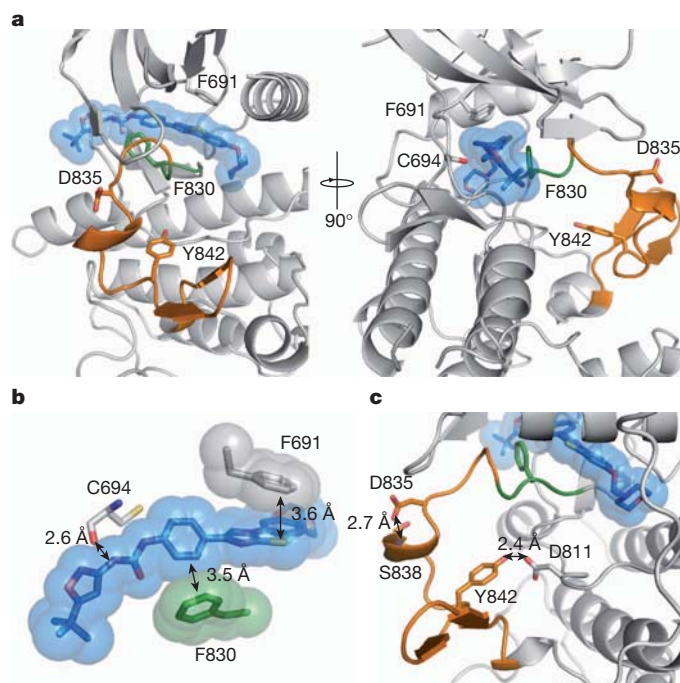


Figure 2 | Modelling of FLT3–AC220 interactions. **a**, Docking model of the AC220-bound FLT3 kinase domain: AC220 (blue); activation loop (orange) and DFG motif (green); amino acid residues that confer AC220 resistance when mutated (F691, D835 and Y842) and that interact with AC220 (F691, C694 and F830) in sticks. **b**, Surface and stick presentation of AC220 and AC220-interacting residues on FLT3: the carbonyl oxygen of C694 interacts with one of the AC220 amide groups; F691 forms a π – π stacking interaction with AC220; F830 interacts with AC220 through a perpendicular aromatic–aromatic interaction. **c**, Structure of the activation loop: residues D835, Y842 and interacting residues on FLT3 depicted in sticks.

was sequenced from PCR-amplified genomic DNA isolated from AC220-resistant clones. Identified drug-resistant mutations were re-engineered into MSCVpuroFLT3-ITD using site-directed mutagenesis and Ba/F3 cell lines were created as detailed in Methods. Cell viability in the presence and absence of drug was assessed using trypan blue exclusion. FLT3 phosphorylation status was determined by western blot analysis of whole cell lysates prepared after 90 min of drug exposure from Ba/F3 cells stably expressing FLT3 mutant isoforms and from transfected 293T cells in the absence of drug. The FLT3 kinase domain was PCR amplified from cDNA derived from blood or bone marrow samples from patients enrolled on the exploratory portion of the phase II trial of AC220 in AML (<http://clinicaltrials.gov/ct2/show/NCT00989261>; identifier NCT00989261) and from normal control bone marrow or mobilized peripheral blood stem cells. PCR products were cloned into *Escherichia coli* and individual clones were sequenced using Sanger sequencing. Alternatively, SMRTBell libraries¹³ were prepared as per the manufacturer's instructions and sequenced on a Pacific Biosciences RS instrument. For details of the computational sequencing analysis, please see Methods. Molecular docking of FLT3-ITD to AC220 was performed using Autodock 4.2 package. Inhibitor binding constants were measured using an active-site-dependent competition binding assay as previously described²⁰. For further details of methods, please see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August 2011; accepted 5 March 2012.

Published online 15 April 2012.

- Thiede, C. *et al.* Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood* **99**, 4326–4335 (2002).
- Lee, B. H. *et al.* FLT3 internal tandem duplication mutations induce myeloproliferative or lymphoid disease in a transgenic mouse model. *Oncogene* **24**, 7882–7892 (2005).
- Shih, L. Y. *et al.* Acquisition of FLT3 or N-ras mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia. *Leukemia* **18**, 466–475 (2004).
- Knapper, S. *et al.* A phase 2 trial of the FLT3 inhibitor lestaurtinib (CEP701) as first-line treatment for older patients with acute myeloid leukemia not considered fit for intensive chemotherapy. *Blood* **108**, 3262–3270 (2006).
- Fischer, T. *et al.* Phase IIB trial of oral Midostaurin (PKC412), the FMS-like tyrosine kinase 3 receptor (FLT3) and multi-targeted kinase inhibitor, in patients with acute myeloid leukemia and high-risk myelodysplastic syndrome with either wild-type or mutated FLT3. *J. Clin. Oncol.* **28**, 4339–4345 (2010).
- Zarrinkar, P. P. *et al.* AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). *Blood* **114**, 2984–2992 (2009).
- Cortes, J. *et al.* in *16th Congress of the European Hematology Association* (Haematologica, 2011).
- Heidel, F. *et al.* Clinical resistance to the kinase inhibitor PKC412 in acute myeloid leukemia by mutation of Asn-676 in the FLT3 tyrosine kinase domain. *Blood* **107**, 293–300 (2006).
- Metzelder, S. *et al.* Compassionate use of sorafenib in FLT3-ITD-positive acute myeloid leukemia: sustained regression before and after allogeneic stem cell transplantation. *Blood* **113**, 6567–6571 (2009).
- Scholl, S. *et al.* Secondary resistance to sorafenib in two patients with acute myeloid leukemia (AML) harboring FLT3-ITD mutations. *Ann. Hematol.* **90**, 473–475 (2011).
- Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/ imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831–843 (2003).
- Shah, N. P. *et al.* Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**, 117–125 (2002).
- Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **38**, e159 (2010).
- Griffith, J. *et al.* The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol. Cell* **13**, 169–178 (2004).
- Levinson, N. M. *et al.* A Src-like inactive conformation in the Abl tyrosine kinase domain. *PLoS Biol.* **4**, e144 (2006).
- Mol, C. D. *et al.* Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **279**, 31655–31663 (2004).
- Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754 (1994).
- Mol, C. D. *et al.* Structure of a c-Kit product complex reveals the basis for kinase transactivation. *J. Biol. Chem.* **278**, 31461–31464 (2003).
- Hubbard, S. R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* **16**, 5572–5581 (1997).
- Wodicka, L. M. *et al.* Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. *Chem. Biol.* **17**, 1241–1249 (2010).
- Burley, S. K. & Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23–28 (1985).
- Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
- Choudhary, C. *et al.* Activation mechanisms of STAT5 by oncogenic Flt3-ITD. *Blood* **110**, 370–374 (2007).
- Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).
- Barbie, D. A. & Deangelo, D. J. Systemic mastocytosis: current classification and novel therapeutic options. *Clin. Adv. Hematol. Oncol.* **4**, 768–775 (2006).
- Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469**, 356–361 (2011).
- Kralovics, R. *et al.* Acquisition of the V617F mutation of JAK2 is a late genetic event in a subset of patients with myeloproliferative disorders. *Blood* **108**, 1377–1380 (2006).
- Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973–977 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Lin for technical assistance. This work was funded by grants from the Leukemia and Lymphoma Society (to C.C.S. and N.P.S.), the Doris Duke Charitable Foundation (to N.P.S.), NCI Leukemia SPORE P50 CA100632-06 (to M.J.L.), NCI R01 CA12886 (to M.J.L.) and the NIH T-32 Molecular Mechanisms of Cancer (to C.C.S.). C.C.S. would like to acknowledge the EHA/ASH Translational Research Training Institute. N.P.S. would like to thank Art and Alison Kern and the Edward S. Ageno family for their support.

Author Contributions C.C.S., Q.W., C.-S.C., K.J.T., A.K., E.E.S. and J.K. designed experiments, performed research, analysed data and wrote the manuscript. N.P.S. designed experiments, analysed data and wrote the manuscript. L.E.D., S.W., J.P.H. and S.S. performed experiments and reviewed the manuscript. P.P.Z. was involved in study design and reviewed the manuscript. A.E.P. and M.J.L. provided reagents, performed research and reviewed the manuscript.

Author Information SMRT sequencing data is deposited online at <http://www.ncbi.nlm.nih.gov/sra> under accession number SRA050226.1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.S. (nshah@medicine.ucsf.edu).

METHODS

DNA constructs, mutagenesis and resistance screen. *FLT3-ITD* cDNA cloned from the MV4;11 cell line (ITD: residues 591–601) into the *HpaI* site of the pMSCVpuro retroviral vector (Clontech) was a gift from Ambit Biosciences and was used as a template for mutagenesis. We used a modified strategy for random mutagenesis previously described¹¹. Briefly, 1 µg of MSCVpuroFLT3-ITD was used to transform the DNA-repair-deficient *Escherichia coli* strain XL-1 Red (Stratagene) and plated on 20 ampicillin-agar bacterial plates. After incubation for 36 h, colonies were collected by scraping, and plasmid DNA was purified by using a plasmid MAXI kit (Qiagen). Subsequently, mutagenized *FLT3-ITD* plasmid stock and Ecopack packaging plasmid were cotransfected into 293T cells grown in DMEM (Invitrogen) containing 10% fetal calf serum (FCS; Omega Scientific) using Lipofectamine 2000 (Invitrogen) as per the manufacturer's protocol. Viral supernatants were collected at 48 h, purified using a 0.44 µm vacuum filter, and used to infect Ba/F3 cells at a 1:100 to 1:300 dilution of viral supernatant to fresh RPMI 1640 (Invitrogen) supplemented with 10% FCS. Alternatively, viral supernatant was aliquoted and frozen. Thawed supernatant was used to infect Ba/F3 cells at a 1:50 dilution. Viral supernatant was diluted with the goal of minimizing multiplicity of infection. For infection, $1-2 \times 10^6$ Ba/F3 cells were resuspended in 3 ml of the diluted viral stock supplemented with recombinant mouse IL-3 (Invitrogen), and 4 µg ml⁻¹ polybrene, plated in each well of a 12-well tissue culture dish and centrifuged at 1,500g in a Beckman Coulter Allegra 6KR centrifuge with a microplate carrier for 90 min at 34 °C. Centrifuged cells were subsequently transferred to a 37 °C incubator overnight. Infected Ba/F3 cells were washed twice with media to remove IL-3 and plated in 3 ml of RPMI medium 1640 at 5×10^5 cells per well of a 6-well dish supplemented with 20% FCS and 1.2% Bacto-agar with 20 nM AC220 (a gift from Ambit Biosciences). After 10–21 days, visible colonies were plucked from agar and expanded in the presence of drug (20 nM AC220).

Sequencing and alignments. Expanded colonies were harvested 7–14 days after isolation from agar, and whole genomic DNA was isolated using the QIAamp kit (Qiagen). The *FLT3* kinase domain was amplified by PCR from whole genomic DNA by using TopTaq DNA polymerase (Qiagen). The primers TK1F (5'-TGCTGTGATACAAATCCCTTGGC-3') and TK2R (5'-TCTCTGCTGAAAGGTCGCCTGTTT-3') were used for kinase domain amplification and subsequent bidirectional sequencing was performed using these primers in addition to TK1R (5'-AGTCTCTCTTCTTCCAGCCTTT-3') and TK2F (5'-GAGAGGCACTCATGTGCAAGTCA-3'). Alignments to the native *FLT3-ITD* sequence were performed using Sequencher software (Gene Codes Corporation). **Generation of mutants.** Mutants isolated in the screen were engineered into pMSCVpuroFLT3-ITD by using the QuikChange mutagenesis kit (Stratagene). In all cases, individual point mutants were confirmed by sequence analysis.

Cell viability and proliferation assays. Stable Ba/F3 lines were generated by using retroviral spinfection with the appropriate mutated plasmid as outlined above, with the exception of the exclusion of polybrene. At 48 h post-infection, puromycin was added to infected cells at a concentration of 4 µg ml⁻¹. Cells were selected in the presence of puromycin for 7–10 days and subsequently IL-3 was washed twice from the cells with media and cells were selected in RPMI medium 1640 plus 10% FCS in the absence of IL-3. Exponentially growing Ba/F3 cells (5×10^4) were plated in each well of a 24-well dish with 1 ml of RPMI 1640 plus 10% FCS containing the appropriate concentration of drug as indicated in triplicate. Cells were allowed to expand for 2 days and were counted by using a Vi-cell XR automated cell viability analyser (Beckman Coulter). The mean number of viable cells at varying concentrations of drug was normalized to the median number of viable cells in the no-drug sample for each mutant. Error bars represent the standard deviation. Numerical IC₅₀ values were generated using nonlinear best-fit regression analysis using Prism 5 software (GraphPad).

For proliferation assays, on day 0, parental Ba/F3 cells and Ba/F3 cells stably expressing *FLT3-ITD* mutant isoforms were plated in triplicate with 1 mL of RPMI 1640 plus 10% FCS at a density of 5×10^4 cells per well in each well of a 24-well dish. Cells were allowed to expand and were counted by using a Vi-cell XR automated cell viability analyser (Beckman Coulter) on days 2, 3, 4, 5, 6 and 7. To maintain exponential growth of cells, 0.5 ml of cells from each well were used for counting on each day and 0.25 ml of cells from the remaining volume were transferred to a new well with 0.75 ml of fresh RPMI plus 10% FCS (including 2 ng ml⁻¹ of IL-3 for parental Ba/F3 cells). Extrapolated cell counts were calculated from the measured count on each day using the appropriate dilution factor (1× on day 2, 4× on day 3, 16× on day 4, and so on). The number of viable cells at each time point was normalized to the starting number of cells for each cell line on day 0 and the mean normalized cell count on each day was calculated. Error bars represent the standard deviation.

Immunoblotting. Exponentially growing Ba/F3 cells stably expressing each mutation along with a native *FLT3-ITD* control were plated in RPMI medium

1640 plus 10% FCS supplemented with kinase inhibitor at the indicated concentration. After a 90-min incubation, the cells were washed in phosphate buffered saline (PBS) and lysed in Cell Extraction Buffer (Invitrogen) supplemented with protease and phosphatase inhibitors. The lysate was clarified by centrifugation and quantified by BCA assay (Thermo Scientific). Protein was subjected to sodium dodecylsulphate polyacrylamide electrophoresis and transferred to nitrocellulose membranes. Immunoblotting was performed using anti-phospho-FLT3, anti-phospho-STAT5, anti-STAT5, anti-phospho-ERK, anti-ERK, anti-phospho-S6, anti-S6, anti-GAPDH (Cell Signaling) and anti-FLT3 S18 antibody (Santa Cruz Biotechnology). Alternatively, 293T cells were plated in 6-cm plates and transfected with MSCVpuroFLT3-ITD plasmid containing *FLT3* mutations of interest using Lipofectamine 2000 (Invitrogen) as per the manufacturer's protocol. After 48 h, cells were washed with PBS, collected, lysed and subjected to western blot analysis as described above.

Competition binding assays. Inhibitor binding constants were measured by using active site-dependent competition binding assays essentially as previously described²⁰. In brief, *FLT3* protein isoforms were labelled with a chimaeric double-stranded DNA tag containing the NFκB binding site (50-GGGAATTCCC-30) fused to an amplicon for qPCR readout, which was added directly to the expression extracts. Binding reactions were assembled by combining DNA-tagged kinase extract, affinity beads loaded with a kinase inhibitor probe molecule, and test compound in 13 binding buffer (PBS, 0.05% Tween 20, 10 mM DTT, 0.1% BSA, 2 mg ml⁻¹ sonicated salmon sperm DNA). Extracts were used directly in binding assays without any enzyme purification steps at a $\geq 10,000$ -fold overall stock dilution (final DNA-tagged enzyme concentration < 0.1 nM). Assays were incubated for 1 h at room temperature (23 °C), which was sufficient to establish equilibrium. Subsequent washing, elution, and qPCR readout steps were as described²⁰.

Patients and *FLT3* kinase domain sequencing analysis. Eight cases of acquired resistance to AC220 were analysed. Patients were enrolled on the exploratory cohort of the phase II clinical trial of AC220 in relapsed or refractory AML at the University of California, San Francisco (UCSF), University of Pennsylvania or Johns Hopkins University. Details of the clinical trials and results are reported elsewhere⁷. All patients were *FLT3-ITD*-positive at enrolment. Samples were collected pre-treatment and at the time of disease progression. Only patients who had achieved morphological clearance of bone marrow blasts to $\leq 5\%$ at best response and subsequently relapsed with an increase in peripheral blood or bone marrow blasts are included in this analysis. The patients in this analysis included all the patients meeting the above criteria at the three participating institutions. All patients gave informed consent according to the Declaration of Helsinki to participate both in the clinical trials and for collection of samples. All research involving human subjects was approved by the relevant Institutional Review Board at each individual participating institution (UCSF, University of Pennsylvania or Johns Hopkins).

For sequencing, frozen Ficoll-purified mononuclear cells obtained from blood or bone marrow were lysed in Trizol (Invitrogen) and RNA was isolated according to the manufacturer's protocol. cDNA was synthesized using Superscript II (Invitrogen) as per the manufacturer's protocol. The *FLT3* kinase domain and adjacent juxtamembrane domain were PCR amplified from cDNA using primers TK1F and TK2R as above. PCR products were cloned using TOPO TA cloning (Invitrogen) and transformed into competent *E. coli*. Individual colonies were plucked, expanded in liquid culture overnight and plasmid DNA for sequencing was isolated using the QIAprep Spin Miniprep kit (Qiagen). Each colony was considered representative of a single mRNA. To minimize contamination from PCR artefact, we sequenced at least 10 and up to 24 *FLT3-ITD*-containing clones from each sample and required that mutations be found in $> 15\%$ of clones. The primers TK1F, TK1R, TK2F and TK2R were used for bidirectional sequencing as above. Alignments with native *FLT3* sequence were performed using Sequencher software (Gene Codes Corporation).

Sample preparation and SMRT sequencing. PCR product containing the *FLT3* kinase domain was generated from patient cDNA as described above using high fidelity DNA polymerase. We prepared PCR products for Pacific Biosciences sequencing using standard commercial kits and reagents (<http://www.pacificbiosciences.com/products/consumables/reagents>) following the manufacturer's instructions. PCR products input amounts ranged from 0.3–3 µg, and we prepared SMRTBell libraries¹³ on the full PCR products without any fragmentation. We sequenced all samples on a Pacific Biosciences RS instrument and recorded sequence for 75 min.

Computational analysis of *FLT3* mutations. We obtained samples from three healthy individuals with no cancer history (two bone marrow, one mobilized peripheral blood stem cells), isolated RNA, made cDNA, amplified the *FLT3* kinase domain, and sequenced following a protocol identical to that used on the AML samples. We used the sequence from Normal Control no. 1 as a control for all process steps between sample acquisition and sequencing. Data from the remaining two normal controls were compared to the Normal Control no. 1

and revealed no significant differences (Supplementary Table 3). We use the circular consensus sequencing (CCS) mode to obtain high accuracy reads for the ~1.4 kb amplicon with PacBio RS. The CCS mode generates reads by combining multiple independent single-pass sequencing reads for individual molecules to correct raw errors and generate a better accuracy consensus (see Supplementary Fig. 2). We report only the CCS reads where the same molecule is sequenced at three or more times, that is, raw read length >4.2 kb for the 1.4 kb amplicon. With the CCS reads, we obtained the sequence of the ~1.4 kb amplicon containing the FLT3 juxtamembrane and kinase domains with up to about 98% to 99% accuracy (see alignment identity for ITD[−] samples in Supplementary Table 2). For each CCS read, we used tandem repeats finder (TRF)²⁹ to identify the ITD sequence. To determine unambiguously whether a read was ITD[−] or ITD⁺ consistently, we used only the CCS reads that included at least the region from the 50-bp 5'-end upstream to the 50-bp 3'-end downstream sequence of the ITD region in the analysis. This allowed us to determine the number of sequences containing the ITD more accurately despite a small percentage of insertion and deletion errors in the CCS reads. Two distinct peaks allowed us to identify ITD[−] versus ITD⁺ CCS reads unambiguously. We found that each sample had only one major ITD as expected, although in some cases the majority ITD differed at relapse compared to pre-treatment. We then passed the ITD⁺ population of the CCS reads to the next stage for codon mutation analysis. A list of the number of total CCS reads identified is listed in Supplementary Table 2. We identified ~200–1,300 CCS reads spanning the whole region between the ITD region and the furthest codon of interest (Y842) for codon analysis per sample.

For codon mutation analysis, we restricted our analysis to the 608, 691, 835 and 842 codons from reference sequence NM_004119 (*Homo sapiens FLT3* mRNA) and then took the frequency of sequences obtained for each of these codons in the PCR amplicon of healthy Normal Control no. 1 and compared that to the frequency of sequences in each AML patient sample. A local quality filter that required exact matching of the codons before and after the codon of interest was used for filtering out low quality codon calls that might be due to sequencing errors. We used the observed frequencies from the control sample for calculating the significance of the observed mutation in the AML patient samples. The *P* value was calculated by comparing the numbers of native codons observed and the alternative codon between the control sample and the AML patient sample with Fisher's exact test on the contingency table^{30,31}. Owing to the potential statistical bias that could arise if the number of observed mutations was small in some cases, or if sequencing error frequencies differed between mutant and reference codon sequences, we only report the mutations using a conservative significance threshold of $P < 1 \times 10^{-5}$. We used a simulation to determine the sensitivity of this analysis to detect a true mutation at a given codon position. Sensitivity in this analysis was determined as a function of three parameters: the number of errors observed in the control sequence, the total number of times that position is sequenced in the control, and the number of times that position is sequenced in the patient sample. For the simulation, we conservatively assumed that all alternative codons seen in the control are actual errors. With this simulation, we estimated that this analysis allows us to detect variants in the patient sample above 3% with high confidence if we get more than 300 observations

of the codon of interest. To refine further our search for mutations underlying relapse in these patients, we considered only those mutations that were in *cis* to an ITD, as defined on being on the same single DNA molecule sequence read. These mutations at both baseline and relapse are listed in Table 2.

Molecular docking. Molecular docking was performed using Autodock 4.2 package³². The FLT-ITD structure (residue 587–947) was prepared from the Protein Data Bank accession 1RJB¹⁴. All bound waters were removed from the protein. The structure was then added for hydrogens, and partial atomic charges were assigned using AutoDockTools (ADT)³². Residues K644, F830, F691 and E661 were selected as flexible residues. To define the flexible residues, we first analysed the crystal structure of imatinib-bound inactive KIT kinase domain. The structure of FLT3 is quite similar to that of KIT, so this comparison helps us identify potential ligand interacting residues in FLT3. In the KIT structure, residue L595, K623, E640, L644, T670, Y672, L799 and F811 are close to the ligand. We thus define the corresponding residues in the FLT3 kinase domain (Y693, F830, F691, K644, E661, L818, L616, M665, V624) as flexible. Our docking studies revealed that the conformations of Y693, L818, L616, M665 and V624 in the docking solutions are largely identical to their conformations in the crystal structure, and so we did not consider these five residues to be flexible in the final calculations.

The coordinates of AC220 were generated using the Dundee PROGRD2 server³³, and its initial conformation was energy minimized by the GROMACS force field. The Gasteiger charges were then assigned to the ligand using ADT. Seven torsion bonds were defined as rotatable during the docking procedure. The ligand was put into the kinase ATP-binding pocket and manually aligned to avoid atom clashes. A three-dimensional grid box (dimensions: 60 × 30 × 60 unit in number of grid points, grid spacing: 0.375 Å) centred at the ligand defining the search space was then created by AutoGrid4.2 (ref. 32). Two hundred runs of Lamarckian Genetic Algorithm were performed to optimize the ligand–protein interactions. The solutions were clustered according to the root mean standard deviation values, and ranked by the binding free energy. Two general poses are observed. The top-ranked pose has an average energy of $-10.32 \text{ kcal mol}^{-1}$ (the lowest energy for this pose is $-10.93 \text{ kcal mol}^{-1}$). The second-ranked pose, which is flipped by 180° with respect to the top-ranked position, has 133 solutions (63%) with an average energy of $-5.82 \text{ kcal mol}^{-1}$ (the lowest energy for this pose is $-6.98 \text{ kcal mol}^{-1}$). Given the gap in the calculated energy, we only picked the lowest one for the purely illustrative purposes of this analysis.

29. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
30. Yates, F. Tests of significance for 2 × 2 contingency tables. *J. R. Stat. Soc. A* **147**, 426–463 (1984).
31. Barnard, G. A. Must clinical trials be large? The interpretation of *p*-values and the combination of test results. *Stat. Med.* **9**, 601–614 (1990).
32. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
33. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).

Selective cortical representation of attended speaker in multi-talker speech perception

Nima Mesgarani¹ & Edward F. Chang¹

Humans possess a remarkable ability to attend to a single speaker's voice in a multi-talker background^{1–3}. How the auditory system manages to extract intelligible speech under such acoustically complex and adverse listening conditions is not known, and, indeed, it is not clear how attended speech is internally represented^{4,5}. Here, using multi-electrode surface recordings from the cortex of subjects engaged in a listening task with two simultaneous speakers, we demonstrate that population responses in non-primary human auditory cortex encode critical features of attended speech: speech spectrograms reconstructed based on cortical responses to the mixture of speakers reveal the salient spectral and temporal features of the attended speaker, as if subjects were listening to that speaker alone. A simple classifier trained solely on examples of single speakers can decode both attended words and speaker identity. We find that task performance is well predicted by a rapid increase in attention-modulated neural selectivity across both single-electrode and population-level cortical responses. These findings demonstrate that the cortical representation of speech does not merely reflect the external acoustic environment, but instead gives rise to the perceptual aspects relevant for the listener's intended goal.

Separating out a speaker of interest from other speakers in a noisy, crowded environment is a perceptual feat that we perform routinely. The ease with which we hear under these conditions belies the intrinsic complexity of this process, known as the cocktail party problem^{1–3,6}; concurrent complex sounds, which are completely mixed upon entering the ear, are re-segregated and selected from within the auditory system. The resulting percept is that we selectively attend to the desired speaker while tuning out the others.

Although previous studies have described neural correlates of masking and selective attention to speech^{4,5,7–9}, fundamental questions remain unanswered regarding the precise nature of speech representation at the juncture where competing signals are resolved. In particular, when attending to a speaker within a mixture, it is unclear what key aspects (for example, spectrotemporal profile, spoken words and speaker identity) are represented in the auditory system and how they compare to representations of that speaker alone; how rapidly a selective neural representation builds up when one attends to a specific speaker; and whether breakdowns in these processes can explain distinct perceptual failures, such as the inability to hear the correct words, or follow the intended speaker.

To answer these questions, we recorded cortical activity from human subjects implanted with customized high-density multi-electrode arrays as part of their clinical work-up for epilepsy surgery¹⁰. Although limited to this clinical setting, these recordings provide simultaneous high spatial and temporal resolution while sampling the population neural activity from the non-primary auditory speech cortex in the posterior superior temporal lobe. We focused our analysis on high gamma (75–150 Hz) local field potentials¹¹, which have been found to correlate well with the tuning of multi-unit spike recordings¹². In humans, the posterior superior temporal gyrus has been heavily implicated in speech perception¹³, and is anatomically defined as the

lateral parabelt auditory cortex (including Brodmann areas 41, 42 and 22)¹⁴.

Subjects listened to speech samples from a corpus commonly used in multi-talker communication research^{15,16}. A typical sentence was “ready tiger go to red two now” where “tiger” is the call sign, and “red two” is the colour–number combination. One male and one female speaker were selected, each speaking the same 12 unique combinations of two call signs (ringo or tiger), three colours (red, blue or green) and three numbers (two, five or seven). Example acoustic spectrograms from two individual speakers are shown in Fig. 1a, b. The two voices differ along several dimensions including pitch (male versus female), spectral profile (different vocal track shapes) and temporal characteristics (speaking rate). Subjects first listened to each of the speakers alone and were able to report the colour and number with 100% accuracy. Subjects then listened to a monaural, simultaneous mixture of the two speakers' phrases with different call signs, colours and numbers. The subjects were instructed to respond by indicating the colour and number spoken by the talker who uttered the target call sign. The target call sign (ringo or tiger) was fixed and shown visually on a monitor during each trial block, which contained 28 different mixture sounds. As the target speaker was changed randomly from trial to trial, the subjects were required to monitor both voices initially (divided attention) to identify the target speaker. The target call sign was switched after each block, turning the previous target speaker in each mixture into a masker. This resulted in two sets of behavioural and neural responses for each identical mixture sound, which differed only in the focus of attention. Subjects reported correct responses in 74.8% of trials.

Figure 1c illustrates the mixture spectrogram and how difficult it is to tell which sound parts belong to one speaker versus the other. The energy for both speakers is distributed broadly across the spectral and temporal domains, with overlap in some areas and isolated sound parts in others, as shown in their difference spectrogram (Fig. 1d; average spectrograms in Supplementary Fig. 1a).

To determine the spectrotemporal encoding of the attended speaker, the method of stimulus reconstruction was used^{17–19} to estimate the speech spectrogram represented by the population neural responses. Reconstructed spectrograms provide an intuitive way to examine how the population neural responses encode the spectrotemporal features of speech, and more importantly, can be compared with the original acoustic spectrograms as well as across attentional conditions. We first calculated the reconstruction filters from a passive listening task using a separate continuous speech corpus (TIMIT²⁰) that consisted of 499 unique short sentences spoken by 402 different speakers. The filters were then fixed and applied to a novel set of population neural responses to the single and attended mixture speech for spectrogram reconstruction.

When listening to a single speaker alone, the reconstructed spectrograms from population neural activity corresponded well to the spectrotemporal features of the original acoustic spectrograms (Fig. 1e, f compared to Fig. 1a, b, respectively), exhibiting fairly precise temporal features and spectral selectivity (for example, correspondence between the high frequency bursts of energy in “tiger” and “two”, in Fig. 1a, b, e, f).

¹Departments of Neurological Surgery and Physiology, UCSF Center for Integrative Neuroscience, University of California, San Francisco, California 94143, USA.

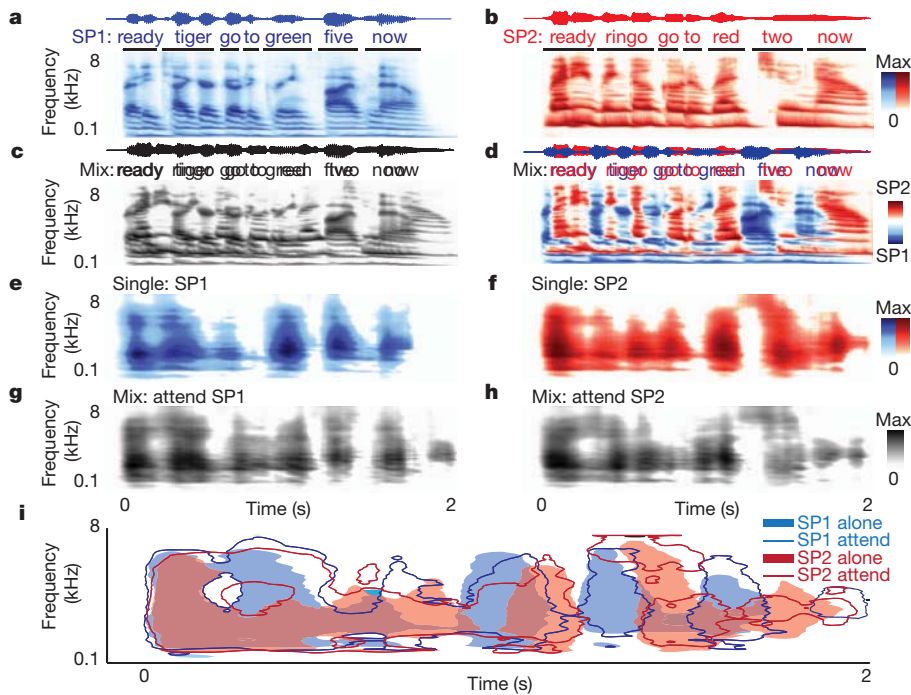


Figure 1 | Acoustic and neural reconstructed spectrograms for speech from a single speaker or a mixture of speakers. **a, b,** Example acoustic waveform and auditory spectrograms of speaker one (male; **a**) and speaker two (female; **b**). **c,** Waveform and spectrogram of the mixture of the two shows highly overlapping energy distributions. **d,** Difference spectrogram highlights the mixture regions where speaker one (blue) or two (red) has more acoustic energy. **e, f,** Neural-population-based stimulus reconstruction of speaker one (**e**) and speaker two (**f**) alone shows similar spectrotemporal features as the original spectrograms in **a** and **b**. **g, h,** The reconstructed spectrograms from the same mixture sound when attending to either speaker one (**g**) or two (**h**) highly resemble the single speaker reconstructions, shown in **e** and **f**, respectively. **i,** Overlay of the spectrogram contours at 50% of maximum energy from the reconstructed spectrograms in **e, f, g** and **h**.

The average and standard deviation of the correlation between reconstructed and original spectrograms over 24 sentences were 0.60 ± 0.034 (0.60 and 0.62 for the examples in Fig. 1e, f). When attending to each of the two speakers, the reconstructed spectrograms from the same speech mixture showed a marked difference depending upon which speaker was attended (Fig. 1g, h). For each pair, the key temporal and spectral features of the target speaker are enhanced relative to the masker speaker (Fig. 1g, h compared to Fig. 1e, f, respectively). To compare directly, the energy contours from these reconstructed spectrograms are overlaid in Fig. 1i. Important spectrotemporal details of the attended speaker were extracted, while the masker speech was effectively suppressed.

Attentional modulation of the neural representation was quantified, separately for correct and error trials, by measuring the correlation of the reconstructed spectrograms from the mixture in two attended conditions with original acoustic spectrograms of the speakers alone (Fig. 2a–d). During correct trials (Fig. 2a, c), we observed a significant shift of average correlation values towards the target speaker representation. During error trials, in contrast, no significant shift was

observed (Fig. 2b, d). Furthermore, the correlations between the reconstructed mixture and the masker speaker were higher than the average intrinsic correlation between randomly chosen original acoustic speech phrases (Fig. 2c, d, dashed lines), revealing a weak presence of the masker speaker in mixture reconstructions, even in correct trials.

The difference in speaking rate of the two speakers, coupled with the stereotyped structure of the carrier phrases, results in specific average temporal modulation profiles for each speaker (average spectrogram for each speaker is shown in Supplementary Fig. 1a, b). To investigate encoding of the distinct spectral profile and characteristic temporal rhythm of the target compared to the masker speaker, we estimated the average difference between reconstructed spectrograms of the two speakers, when presented alone and in the attended mixture (Fig. 2e, f). The comparison between the two average difference reconstructed spectrograms reveals enhanced encoding of both temporal and spectral aspects of the attended speaker (Supplementary Fig. 1c, d). To study the time course of attention-induced modulation of reconstructed mixture spectrograms towards the attended speaker, we

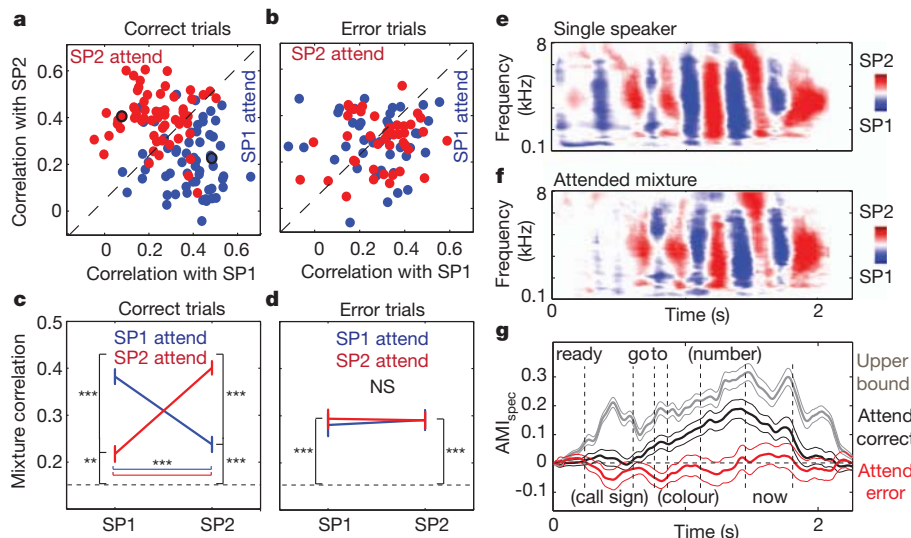


Figure 2 | Quantifying the attentional modulation of neural responses. **a, b,** Correlation coefficients of reconstructed mixture spectrograms under attentional control and the corresponding single speaker original spectrograms in correct and error trials (examples in Fig. 1g, h shown with black outline). **c, d,** Mean and standard error of correlation values for correct and error trials (28 mixtures). The dashed line corresponds to the average intrinsic correlation between randomly chosen original speech phrases. Brackets indicate pairwise statistical comparisons. NS, not significant. **e, f,** Average difference reconstructed spectrograms of speakers one and two from responses to single speaker (**e**) and attended mixture (**f**). **g,** Time course of average and standard error of AMI_{spec} of 28 mixtures for correct (black) and error (red) trials. Grey curve shows the upper bound of AMI_{spec} .

calculated an attentional modulation index (AMI_{spec}), using a sliding window of 250 ms throughout the trial duration:

$$AMI_{\text{spec}} = \text{Corr}(SP1_{\text{spec}}, SP1_{\text{attend}}) - \text{Corr}(SP1_{\text{spec}}, SP2_{\text{attend}}) + \text{Corr}(SP2_{\text{spec}}, SP2_{\text{attend}}) - \text{Corr}(SP2_{\text{spec}}, SP1_{\text{attend}}) \quad (1)$$

where $SP1_{\text{spec}}$ and $SP2_{\text{spec}}$ are the original acoustic spectrograms of speakers one and two, respectively, and $SP1_{\text{attend}}$ and $SP2_{\text{attend}}$ are the spectrograms reconstructed from neural responses to the mixture with attended targets, speaker one and two, respectively. Positive values of this index reflect shifts towards the target, negative values reflect shifts to the masker representation, and values around zero reflect no shift ($AMI_{\text{spec}} = 0.58$ for the example in Fig. 1). An upper bound for the AMI_{spec} was calculated by assuming that attention, at best, restores the single speaker reconstructions of the target speaker (replacing $SP1_{\text{attend}}$ and $SP2_{\text{attend}}$ in equation (1) with $SP1_{\text{alone}}$ and $SP2_{\text{alone}}$; Fig. 2g, grey line). The AMI_{spec} from the mixture was first estimated from correct trials (Fig. 2g, black line), and could resolve the time point at which the reconstructed spectrograms were modulated by attention. After the end of the call sign, which cues the speaker that should be attended, a rapid positive shift in the AMI_{spec} was observed, implying the enhanced representation of the target speaker. In error trials, this effect shows a bias towards the masker speaker, which, in contrast, occurred far earlier in the time course. The neural response shift towards the masker, which occurs as early as the call sign, suggests that listeners had prematurely attended to the wrong speaker during those error trials.

Although the reconstruction analyses showed clear attention-based spectrotemporal modulation, we wanted to determine explicitly whether the attended speech in a mixture could be decoded from a model of a single speaker. A regularized linear classifier²¹ was trained on neural responses to the single speakers and then used to decode both the spoken words and speaker identity of the attended speech mixture. To keep the chance performance at 50% across all comparisons, classification results were limited only to the choices that were present in each mixture. For correct trials, the colour and number of the attended speech were decoded with high accuracy (77.2% and 80.2%, $P < 10 \times 10^{-4}$, t -test; Fig. 3a). However, the decoding performance during error trials was significantly below chance (30.0%, 30.1%, $P < 10 \times 10^{-4}$, t -test; Fig. 3b), indicating a systematic bias towards decoding the words of the masker speaker. In addition, for correct trials, the call sign was classified at chance performance (Fig. 3a). However, for incorrect trials the classifier detected the masker call sign significantly more often than the target call sign (34.1%, $P < 10 \times 10^{-4}$, t -test; Fig. 3b), which again shows errors due to an early selection of the masker (incorrect) speaker.

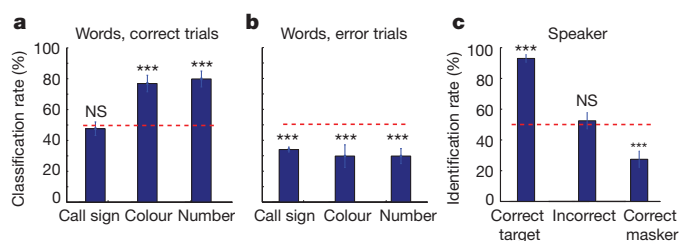


Figure 3 | Decoding spoken words and the identity of the attended speaker. **a**, Classification rate and standard deviation for spoken words (call sign, colour and number) of the attended speaker from the neural responses to the 28 mixtures. Classifiers were trained on single speaker examples only. Colour and number of the attended speech are decoded with high accuracy (77.2% and 80.2%, $P < 10 \times 10^{-4}$, t -test) in correct trials, but not the call sign (48.0%, not significant (NS), t -test). **b**, In error trials, the classifier showed a systematic bias towards the words of the masker speaker (34.1%, 30.0%, 30.1%, $P < 10 \times 10^{-4}$, t -test). **c**, Attended speaker identification rate and standard deviation in correct for target, incorrect (for both target and masker), and correct for masker trials.

For the speaker identification analyses, we divided the behavioural error types into two subsets. The first type occurred when the reported colour-number combination was incorrect for either speaker ('incorrect'; 16.5% of trials). The second type occurred when subjects reported the correct colour-number for the masker instead of the target speaker ('correct for masker'; 8.6% of trials).

In correct trials, the classifier identified the target speaker 93.0% of the time ($P < 10 \times 10^{-4}$, t -test; Fig. 3c). During incorrect trials, the classifier performance was at chance. However, during correct for masker trials, the classifier identified the masker rather than the target speaker (27.3%; $P < 10 \times 10^{-4}$, t -test; Fig. 3c). These classification results confirm the observed restoration seen in spectrotemporal reconstruction, without necessarily assuming a linear relationship between the neural responses and the stimulus. Furthermore, they extend recent findings using similar methods to decode speech sounds presented in isolation²² to full words and sentences under complex listening conditions.

We next asked whether the observed robust encoding of attended speech results as an emergent property of the distributed population activity or is driven by a few spatially discrete sites. The cortical regions with reliable evoked responses to speech stimuli were found using a t -test between neural responses during speech and silence ($P < 0.01$), and were confined to the posterior superior and middle temporal gyri (Fig. 4a). An example of the attentional response modulation at a single electrode is shown in Fig. 4b–d. The spectrotemporal receptive field (STRF, estimated using the <http://www.strflab.berkeley.edu> package) of this electrode in passive listening to speech (TIMIT²⁰) showed a strong preference for high frequency sounds (Fig. 4b) (STRFs for all electrodes of one subject are provided in Supplementary Fig. 2b). This tuning was also evident in the increased neural response at this electrode (Fig. 4d, dashed lines) to each of the single speakers' high frequency sound components (circled in Fig. 4c, responses are delayed about 120 ms from the stimulus). However, the responses to the same speech mixture sound (Fig. 4d, solid lines) were significantly modulated by attention. The responses to high frequency components were enhanced for the attended speaker, but suppressed for similar sounds in the masker speaker (Fig. 4d, solid lines compared to dashed lines). This highly modulated yet fixed feature selectivity probably contributes to the constancy of the single speaker representation observed in our previous analyses. To quantify this effect for each individual electrode, we measured the correlation between the neural responses to the attended mixture and to those of the speakers in isolation (AMI_{elec} , equation (2) in Methods). We found a varying degree of bias towards the attended speaker distributed across the population (Supplementary Fig. 3d; $AMI_{\text{elec}} = 0.28$ for the example

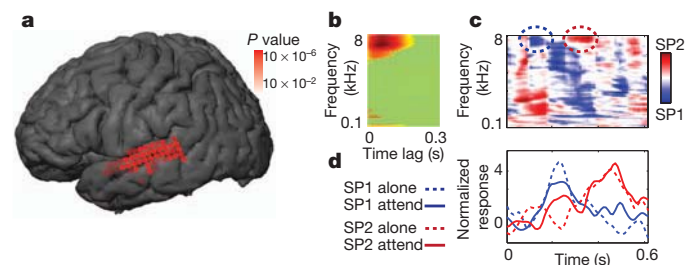


Figure 4 | Attentional modulation of individual electrode sites. **a**, Electrodes picking up a significant difference between responses to silence and speech sounds ($P < 0.01$, t -test). **b**, STRF of this representative electrode site shows a preference for high frequency sounds. **c**, Mixture difference spectrogram for a selected duration containing a high frequency component for each speaker (circled). **d**, The electrode shows an increased response to high frequency sounds of single speakers (dashed lines, peak neural response is delayed by about 120 ms). However, the neural response to the same mixture sound in two attention conditions (solid lines) showed an enhanced response to high frequency sounds only for the target, but with responses for similar sounds in the masker speaker suppressed.

in Fig. 4), which gradually builds up after the end of the call sign (Supplementary Fig. 3e). We did not observe any particular anatomical pattern for the attentional modulation across sites (Supplementary Fig. 3f). Rather, it appeared to be distributed over responsive sites, consistent with previous findings of higher-order sound processing²³.

In summary, we demonstrate that the human auditory system restores the representation of the attended speaker while suppressing irrelevant competing speech. Speech restoration occurs at a level where neural responses still show precise phase-locking to spectrotemporal features of speech. Population responses revealed the emergent representation of speech extracted from a mixture, including the moment-by-moment allocation of attentional focus.

These results have implications for models of auditory scene analysis. In agreement with recent studies, the cortical representation of speech in the posterior temporal lobe does not merely reflect the acoustical properties of the stimulus, but instead relates strongly to the perceived aspects of speech¹⁰. Although the exact mechanisms are not fully known, multiple processes in addition to attention are likely to enable this high-order auditory processing, including grouping of predictable regularities in speech acoustics²⁴, feature binding^{3,25} and phonemic restoration²⁶. Conversely, behavioural errors seem to result from degradation of the neural representation, a direct result of inherent sensory interference such as energetic masking¹⁶ (Supplementary Fig. 3g, h) and/or the allocation of attention²⁷.

In speech, the end result represented in the posterior temporal lobe appears to be unaffected by perceptually irrelevant sounds, which is ideal for subsequent linguistic and cognitive processing. Following one speaker in the presence of another can be trivial for a normal human listener, but remains a major challenge for state-of-the-art automatic speech recognition algorithms²⁸. Understanding how the brain solves this problem may inspire more efficient and generalizable solutions than current engineering approaches²⁹. It will also shed light on how these processes become impaired during ageing and in disorders of speech perception in real-world hearing conditions⁷.

METHODS SUMMARY

Three human subjects with normal hearing underwent the placement of a subdural electrode array as part of their clinical treatment for epilepsy. We used speech samples from a publicly available database called Coordinate Response Measure (CRM¹⁵). One male and one female speaker were selected with two call signs (ringo and tiger), three colours (red, blue or green) and three numbers (two, five or seven). We generated 12 unique combinations of call sign, colour and number per speaker (total of 24 single speaker phrases) and 28 mixture speech samples by selecting from combinations of the 24 single speaker sentences (0 dB target-to-masker ratio). Speech sounds were presented monaurally from a loud speaker. We used stimulus reconstruction^{17–19} to map the population electrocorticographic response to the spectrogram of the speech stimulus. Reconstruction filters were estimated from neural responses to a separate speech corpus (TIMIT²⁰). Test speakers were not used in the estimation of filters. For word and speaker decoding analysis, a regularized linear classifier²¹ was trained on neural responses of the single speakers and then used to decode the spoken words and speaker identity of the attended speech mixture.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 August 2011; accepted 5 March 2012.

Published online 18 April 2012.

1. Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* **25**, 975–979 (1953).
2. Shinn-Cunningham, B. G. Object-based auditory and visual attention. *Trends Cogn. Sci.* **12**, 182–186 (2008).

3. Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, 1994).
4. Kerlin, J., Shahin, A. & Miller, L. Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* **30**, 620–628 (2010).
5. Besle, J. *et al.* Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci.* **31**, 3176–3185 (2011).
6. Bee, M. & Micheyl, C. The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J. Comparative Psychol.* **122**, 235–252 (2008).
7. Shinn-Cunningham, B. G. & Best, V. Selective attention in normal and impaired hearing. *Trends Amplif.* **12**, 283–299 (2008).
8. Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P. & Wise, R. J. S. The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *J. Acoust. Soc. Am.* **125**, 1737–1743 (2009).
9. Elhilali, M., Xiang, J., Shamma, S. A. & Simon, J. Z. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol.* **7**, e1000129 (2009).
10. Chang, E. F. *et al.* Categorical speech representation in human superior temporal gyrus. *Nature Neurosci.* **13**, 1428–1432 (2010).
11. Crone, N. E., Boatman, D., Gordon, B. & Hao, L. Induced electrocorticographic gamma activity during auditory perception. *Clin. Neurophysiol.* **112**, 565–582 (2001).
12. Steinschneider, M., Fishman, Y. I. & Arezzo, J. C. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb. Cortex* **18**, 610–625 (2008).
13. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
14. Hackett, T. A. Information flow in the auditory cortical network. *Hear. Res.* **271**, 133–146 (2011).
15. Bolia, R. S., Nelson, W. T., Ericson, M. A. & Simpson, B. D. A speech corpus for multitaler communications research. *J. Acoust. Soc. Am.* **107**, 1065–1066 (2000).
16. Brungart, D. S. Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* **109**, 1101–1109 (2001).
17. Mesgarani, N., David, S. V., Fritz, J. B. & Shamma, S. A. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* **102**, 3329–3339 (2009).
18. Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. & Warland, D. Reading a neural code. *Science* **252**, 1854–1857 (1991).
19. Pasley, B. N. *et al.* Reconstructing speech from human auditory cortex. *PLoS Biol.* **10**, e1001251 (2012).
20. Garofolo, J. S. *et al.* *TIMIT Acoustic-Phonetic Continuous Speech Corpus* (Linguistic Data Consortium, 1993).
21. Rifkin, R., Yeo, G. & Poggio, T. Regularized least-squares classification. *Nato Science Series Sub Series III Computer and Systems Sciences* **190**, 131–154 (2003).
22. Formisano, E., De Martino, F., Bonte, M. & Goebel, R. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* **322**, 970–973 (2008).
23. Staeren, N., Renvall, H., De Martino, F., Goebel, R. & Formisano, E. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* **19**, 498–502 (2009).
24. Shamma, S. A., Elhilali, M. & Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **34**, 114–123 (2010).
25. Darwin, C. J. Auditory grouping. *Trends Cogn. Sci.* **1**, 327–333 (1997).
26. Warren, R. M. Perceptual restoration of missing speech sounds. *Science* **167**, 392–393 (1970).
27. Kidd, G. Jr, Arbogast, T. L., Mason, C. R. & Gallun, F. J. The advantage of knowing where to listen. *J. Acoust. Soc. Am.* **118**, 3804–3815 (2005).
28. Shen, W., Olive, J. & Jones, D. Two protocols comparing human and machine phonetic discrimination performance in conversational speech. *INTERSPEECH* 1630–1633 (2008).
29. Cooke, M., Hershey, J. R. & Rennie, S. J. Monaural speech separation and recognition challenge. *Comput. Speech Lang.* **24**, 1–15 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank A. Ren for technical help, and C. Micheyl, S. Shamma and C. Schreiner for critical discussion and reading of the manuscript. E.F.C. was funded by National Institutes of Health grants R00-NS065120, DP2-OD00862, R01-DC012379, and the Ester A. and Joseph Klingenstein Foundation.

Author Contributions N.M. and E.F.C. designed the experiment, collected the data, evaluated results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.F.C. (changed@neurosurg.ucsf.edu).

METHODS

The experimental protocol was approved by the Committee for Human Research at the University of California, San Francisco.

Subjects. Three human subjects underwent the placement of a high-density subdural electrode array (4 mm pitch) over the language-dominant hemisphere as part of routine clinical treatment for epilepsy. Subjects gave their written informed consent before surgery. All subjects had self-reported normal hearing and underwent neuropsychological language testing (including the Boston naming and verbal fluency tests) and were found to be normal. The intracarotid sodium amobarbital (Wada) test was used for language dominance assessment. The electrodes in the study were located over the posterior dorsolateral temporal lobe. The location and corresponding spectrotemporal receptive fields of all the included electrodes for a subject are shown in Supplementary Fig. 2.

Data acquisition and pre-processing. The electrocorticography signal was recorded with a multichannel amplifier optically connected to a digital signal processor (TuckerDavis Technologies). Each channel time series was visually and quantitatively inspected for artefacts or excessive noise. The data were then segmented with a 100 ms pre-stimulus baseline and a 400 ms post-stimulus interval. The common mode signal was estimated using principal component analysis with channels as repetitions and was removed from each channel time series using vector projection.

Task design and behavioural testing. We used speech samples from a publicly available database called Coordinate Response Measure (CRM¹⁵) containing sentences in the form “ready (call sign) go to (colour) (number) now”. One male and one female speaker (speakers one and five in CRM corpus) were selected with two call signs (ringo and tiger), three colours (blue (B), red (R) or green (G)) and three numbers (two, five or seven). For each of the two call signs, we generated six colour-number combinations (B2, B5, R2, R7, G5, G7), resulting in 12 different phrases. We chose the same phrases for each of the two speakers, resulting in 24 single speaker sentences. We then produced 28 unique mixture speech samples by selecting from combinations of the 24 single speaker sentences at 0 dB target-to-masker ratio. Each mixture sample was chosen such that there was no overlap between call signs, colours or the numbers of the two phrases. In addition, each speaker had the same number of call signs (ringo or tiger) in each trial block. The sounds were presented monaurally from a loudspeaker connected to a laptop, which was also used to collect subjects’ responses through a customized graphical user interface. Each trial block consisted of 28 trials and the target call sign was fixed for each block. The target call sign was displayed visually before and during the trial block. Subjects first listened to each of the speakers alone and were able to report the colour and number with 100% accuracy. Subjects then listened to a monaural, simultaneous mixture of the two speakers’ phrases with different call signs, colours and numbers. The subjects were instructed to respond by indicating the colour and number spoken by the talker who uttered the target call sign. The target speaker changed from trial to trial pseudo-randomly, requiring the subjects to initially monitor both speakers until they detect the target call sign. After each trial block, the target call sign was changed, switching the role of target and masker speakers in each mixture sound.

Electrode selection. The cortical sites on the superior and middle temporal gyri with reliable evoked responses to speech stimuli were selected for all the subsequent analysis. Our inclusion criteria consisted of a *t*-test between responses to randomly selected time frames during passive speech presentation (TIMIT) and in silence ($P < 0.01$, resulting in 83, 92 and 102 electrodes for subjects one to three. One example subject is shown in Supplementary Fig. 2a). Solely for visualization, we also estimated the STRFs of these selected sites from passive

listening to TIMIT using normalized reverse correlation algorithm (STRFLab software package, <http://www.strflab.berkeley.edu>; Supplementary Fig. 2b). Correlation histogram of STRF predictions for all 275 electrode sites is shown in Supplementary Fig. 1c.

Stimulus reconstruction. We used stimulus reconstruction to map the population neural responses to the spectrogram of the speech stimulus^{17–19}. Reconstruction filters were estimated from neural responses to a separate speech corpus (TIMIT²⁰) containing a total of 499 unique short sentences from 402 different speakers. Filters were obtained using normalized reverse correlation to minimize the mean squared error of the reconstructed spectrograms¹⁷ with filter time lags from -420 to 0 ms (causal filters). The filters were then fixed in all subsequent conditions and were applied to the neural responses to CRM samples. Neither of the speakers or phrases in the CRM data set was used in estimation of the filters. The output of the reconstruction algorithm was further processed with a band-pass filter applied to each frequency channel of reconstructed spectrograms to remove the baseline. All the processing steps for stimulus reconstruction were identical in all conditions (single and mixture speakers).

AMI. To quantify the change in similarity between the representation of single and attended speaker in mixture speech, we defined the AMI_{spec} in equation (1). The stereotypical format of the CRM phrases results in an intrinsic correlation between the neural responses to different sentences, particularly at the beginning (“ready”) and middle of the carrier phrase (“go to”), which results in reduced possible AMI_{spec} values for these segments. To estimate an upper bound for unbiased comparison, AMI_{spec} was calculated where the representation of an attended speaker in a mixture is ideally assumed to be identical to the representation of that speaker when presented alone; therefore, replacing SP_{attend} in equation (1) with the reconstructed spectrogram of single speaker SP_{alone} . The upper bound peaks at the call sign, colour and number where different phrases are most dissimilar. The overall increase in the upper bound is due to the progressive asynchrony between the two speakers.

The same statistics can be used to estimate the AMI of an individual electrode site by calculating the correlation values between the neural response of that site to attended mixture and single speaker presentations:

$$AMI_{\text{elec}} = \text{Corr}(R\text{-}SP1_{\text{alone}}, R\text{-}SP1_{\text{attend}}) - \text{Corr}(R\text{-}SP1_{\text{alone}}, R\text{-}SP2_{\text{attend}}) + \text{Corr}(R\text{-}SP2_{\text{alone}}, R\text{-}SP2_{\text{attend}}) - \text{Corr}(R\text{-}SP2_{\text{alone}}, R\text{-}SP1_{\text{attend}}) \quad (2)$$

where $R\text{-}SP1_{\text{alone}}$ and $R\text{-}SP2_{\text{alone}}$ are the responses of an electrode to speakers one and two alone, respectively, and $R\text{-}SP1_{\text{attend}}$ and $R\text{-}SP2_{\text{attend}}$ are the responses of the same electrode to the mixture of the two when the attended target is speaker one and two, respectively.

Classification of spoken words and speaker identity. A linear-frame-based regularized-least-square classifier²¹ was used to investigate the discriminability of the spoken words and speaker identity from electrocorticographic responses. Two binary classifiers were trained to classify the call sign and speaker identity, and two separate three-way classifiers were used for colour and for number classification. Classifiers were trained only on the neural responses of single speakers (24 sentences) and tested on the mixtures. The classifiers produced a linear weighted sum of the neural responses at each time instance and the classifier that produced the maximum average output over the duration of words was chosen as classification result. The classifier decision was limited to only the colours and numbers that occurred in each mixture, therefore resulting in same 50% chance performance in all cases.

SOLID-STATE PHYSICS

Electrons do the split

Interacting electrons that are confined to move in a one-dimensional structure do not simply jam together like cars in rush hour. Inelastic X-ray scattering shows that the electrons act as if they split into separate fractional entities.

RALPH CLAESSEN

In physics, a phenomenon is often best explained by reducing it to its simplest version. For example, to understand the quantum-mechanical motion of non-interacting electrons in a crystalline solid, the case of a one-dimensional crystal lattice suffices to introduce the concept of electronic band structure, which describes the range of energies that the electrons may have in the solid. But when it comes to interacting electrons, this approach fails. Coulomb repulsion between any two electrons is much stronger in one-dimensional solids than in their higher-dimensional counterparts, and many-body effects emerge that lead to an apparent fractionalization of the electron^{1–5}. With this fractionalization, the electron's spin and charge seem to form separate quasiparticles — spinons and holons, respectively — that move independently of each other and have different velocities. In a paper published on *Nature's* website today, Schlappa *et al.*⁶ describe an experiment that takes the idea of electronic fractionalization a step further by revealing that electrons can split into a third form of quasiparticle, the orbiton.

Schlappa and colleagues observe orbitons in a copper oxide (Sr_2CuO_3) containing a one-dimensional chain of copper oxide building blocks (CuO_3), in which the valence electrons reside in the 3d electronic shell of the central copper atoms (Fig. 1). In a free atom there are five 3d orbitals, which are energetically degenerate — that is, they all have the same energy. However, in a CuO_3 unit, the electrical field of the oxygen ions around each copper atom lifts this degeneracy, generating an intra-shell energy spectrum. In the chain's ground state, the 3d electrons in all of the CuO_3 units are in the same, lowest-energy 3d state, with their spin orientation alternating between neighbouring units as a result of antiferromagnetic interactions.

In such a system, electrons can be locally excited between the energetically different 3d states, leaving one of the CuO_3 units in a state known as crystal-field excitation. To achieve this excitation, Schlappa *et al.* shone X-rays on the CuO_3 chain using a technique called resonant inelastic X-ray scattering. The method is similar to conventional optical

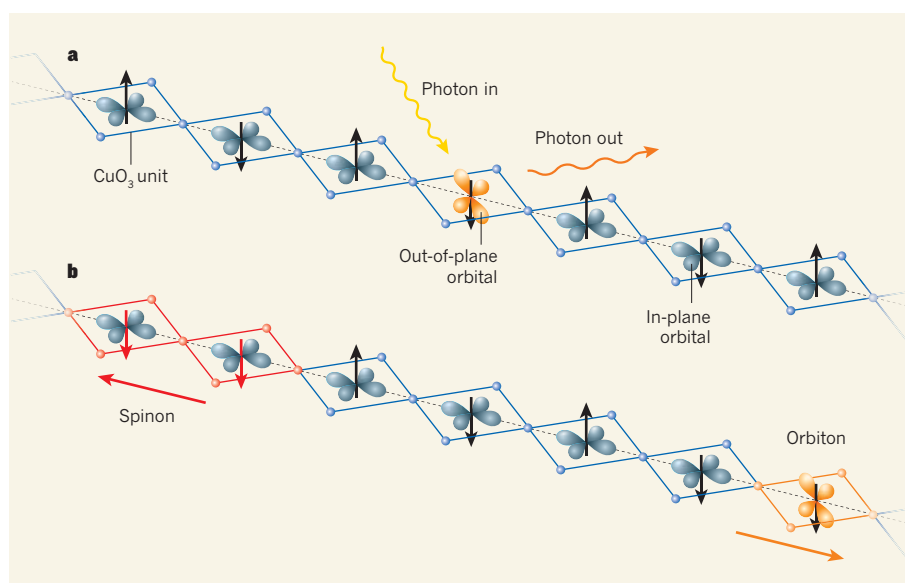


Figure 1 | Electron fractionalization in a one-dimensional structure. **a**, Schlappa and colleagues' system⁶ consists of a chain of copper oxide units (CuO_3). In the ground state, all units contain 3d electrons in the same in-plane orbital, and the direction of the local 3d spins (arrows) alternates between neighbouring units. Shining X-ray photons on the chain excites the electrons in one of the units into an out-of-plane orbital of higher energy. **b**, The authors show that the X-ray irradiation leads to an immediate splitting of the excitation into two separate quasiparticles of different velocities: a spinon, which is associated with a local perturbation in the spin arrangement, and an orbiton, which carries the orbital excitation.

Raman scattering, but owing to the short wavelengths of X-rays, it provides information not only about a sample's electronic excitations, but also about the excitations' spatial dynamics.

Traditionally, crystal-field excitations in 3d-metal compounds have been viewed as localized objects that stay fixed to the atomic site at which they were generated. But Schlappa and colleagues⁶ provide clear evidence that, in one dimension, such entities break up into two independent, mobile fragments immediately after excitation. These fragments are the spinon, which is a local disturbance of the spin arrangement of the electron ensemble, and the orbiton, a collective response of all 3d electrons in the chain, which carries the 3d intra-shell excitation. The authors find that the separation of the spinon and the orbiton and their ensuing dynamical behaviour are in excellent agreement with theoretical expectations.

The concept of orbitons as collective orbital excitations in 3d-metal compounds, which is

similar to that of spin waves in magnetically ordered systems, was introduced more than a decade ago⁷. However, for a long time their propagating character escaped unambiguous experimental verification, because previous work⁸ had been directed at materials of higher dimensions than the one-dimensional system studied here. In such systems, intricate coupling of the initial intra-shell 3d excitation to magnetic excitations tends to immobilize the orbitons, preventing their observation as propagating entities. The spinon–orbiton separation reported by Schlappa *et al.* not only establishes a new aspect of one-dimensional electron fractionalization, but also provides the first clear observation of moving orbitons in a solid.

It should be noted that this observation has become possible only through the enormous advances made in resonant inelastic X-ray scattering during the past few years, notably through pioneering work at the Swiss Light Source facility at the Paul Scherrer Institute

in Villigen, Switzerland, where the authors' experiments⁶ were conducted. Schlappa and colleagues' results thus highlight the increasingly crucial role of resonant inelastic X-ray scattering in the study of electronic excitations, just as inelastic neutron scattering is the method of choice in the study of magnetic and lattice excitations⁸.

Finally, are there any practical implications for spinon–orbiton separation? As the microelectronics industry moves towards ever-increasing miniaturization, electronic devices and the conducting connections between them

will eventually reach quantum-mechanical limits. When the diameter of the connections reaches atomic dimensions, the transport of charge in them — and also of spin and heat — will no longer follow the laws for macroscopic (three-dimensional) conductors, and one-dimensional electron fractionalization will become relevant. Whether such fractionalization can be used for new device functionalities in quantum computing or spin electronics remains to be seen. ■

Ralph Claessen is at the *Physikalisches*

*Institut, Universität Würzburg,
D-97074 Würzburg, Germany.
e-mail: claessen@physik.uni-wuerzburg.de*

1. Voit, J. *Rep. Prog. Phys.* **58**, 977–1116 (1995).
2. Kim, C. *et al. Phys. Rev. Lett.* **77**, 4054–4057 (1996).
3. Claessen, R. *et al. Phys. Rev. Lett.* **88**, 096402 (2002).
4. Auslaender, O. M. *et al. Science* **308**, 88–92 (2005).
5. Kim, B. J. *et al. Nature Phys.* **2**, 397–401 (2006).
6. Schlappa, J. *et al. Nature* <http://dx.doi.org/10.1038/nature10974> (2012).
7. Saitoh, E. *et al. Nature* **410**, 180–183 (2001).
8. Ament, L. J. P., van Veenendaal, M., Devereaux, T. P., Hill, J. P. & van den Brink, J. *Rev. Mod. Phys.* **83**, 705–767 (2011).

In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes

Li Qian^{1,2,3}, Yu Huang^{1,2,3}, C. Ian Spencer^{1,2,3}, Amy Foley^{1,2,3}, Vasanth Vedantham^{1,4,5}, Lei Liu^{1,2,3}, Simon J. Conway⁶, Ji-dong Fu^{1,2,3} & Deepak Srivastava^{1,2,3}

The reprogramming of adult cells into pluripotent cells or directly into alternative adult cell types holds great promise for regenerative medicine. We reported previously that cardiac fibroblasts, which represent 50% of the cells in the mammalian heart, can be directly reprogrammed to adult cardiomyocyte-like cells *in vitro* by the addition of Gata4, Mef2c and Tbx5 (GMT). Here we use genetic lineage tracing to show that resident non-myocytes in the murine heart can be reprogrammed into cardiomyocyte-like cells *in vivo* by local delivery of GMT after coronary ligation. Induced cardiomyocytes became binucleate, assembled sarcomeres and had cardiomyocyte-like gene expression. Analysis of single cells revealed ventricular cardiomyocyte-like action potentials, beating upon electrical stimulation, and evidence of electrical coupling. *In vivo* delivery of GMT decreased infarct size and modestly attenuated cardiac dysfunction up to 3 months after coronary ligation. Delivery of the pro-angiogenic and fibroblast-activating peptide, thymosin β 4, along with GMT, resulted in further improvements in scar area and cardiac function. These findings demonstrate that cardiac fibroblasts can be reprogrammed into cardiomyocyte-like cells in their native environment for potential regenerative purposes.

Heart failure affects over 14 million people worldwide and is a leading cause of death in adults and in children. Because postnatal cardiomyocytes (CMs) have little or no regenerative capacity, therapies are limited at present. The introduction of exogenous stem-cell-derived CMs holds promise, but also challenges, including delivery, integration, rejection and cellular maturation^{1–3}. Reprogramming adult fibroblasts into induced pluripotent stem cells (iPSCs) that are similar to embryonic stem cells addresses some issues^{4–6}, but others, including efficient directed differentiation into CMs and effective delivery, remain.

A new generation of reprogramming technology involves trans-differentiating one adult somatic cell type directly into another^{7–11}. We reported direct reprogramming of fibroblasts into CM-like cells *in vitro* by expressing three transcription factors: Gata4, Mef2c and Tbx5 (GMT)⁷. As observed in reprogramming to iPSCs, the percentage of fibroblast cells fully reprogrammed to beating CMs *in vitro* was small, but far more were partially reprogrammed, much like pre-iPSCs that can become fully pluripotent with additional stimuli¹². We posited that cardiac fibroblasts may reprogram more fully *in vivo* in their native environment, which might promote survival, maturation, and coupling with neighbouring cells. If so, the vast pool of cardiac fibroblasts in the heart could serve as an endogenous source of new CMs for regenerative therapy.

Retroviral delivery of GMT *in vivo*

We used a retroviral system to express GMT, and/or dsRed as a marker, in the hearts of 2-month-old male mice by direct intramyocardial injection. After 2 days, transverse sections of the injected area were prepared and co-stained for dsRed, α -actinin (a CM marker) and vimentin (enriched in fibroblasts). No markers are uniquely specific for cardiac fibroblasts, but fibroblasts are known to express vimentin and the surface markers Thy1 and DDR2 (ref. 13). At baseline, it was difficult to detect α -actinin- or vimentin-positive cells that also

expressed dsRed, suggesting minimal viral uptake, and consistent with the observation that retroviruses only infect actively dividing cells¹⁴.

Fibroblasts are embryologically distinct from CMs in their origin¹⁵, and following myocardial infarction (MI) become activated, migrate to the injury site, and proliferate^{16,17}. We induced cardiac injury by coronary artery ligation and injected dsRed retrovirus into the myocardium bordering the infarct zone. Whereas cells co-expressing dsRed and α -actinin were still undetectable, many vimentin-positive cells were also positive for dsRed (Supplementary Fig. 1). By fluorescence-activated cell sorting (FACS), over 4% of cells ($98,238 \pm 5,523$) from the left ventricle of injected hearts were dsRed⁺ Thy1⁺ 2 days after injury, suggesting successful delivery of virus into cardiac fibroblasts and possibly other non-myocytes upon injury (Fig. 1a, b). By quantitative polymerase chain reaction (qPCR), dsRed⁺ Thy1⁺-sorted cells expressed about 60-fold more GMT than dsRed⁺ Thy1⁺ cells, and 6–8-fold more than endogenous CMs (Fig. 1c). A similar number of dsRed⁺ Thy1⁺ cells represented other non-myocyte cell types. Endothelial cells (PECAM⁺) and some perivascular cells (NG2⁺; also known as Cspg4) were also transduced by the retrovirus, but haematopoietic (CD34⁺) and pericardial (WT1⁺) cells were not (Supplementary Fig. 1).

Reprogramming into induced cardiomyocytes

To determine whether new cardiomyocytes could be created *in vivo* from cells other than post-mitotic CMs, we used lineage-tracing experiments to track the origin of putative induced cardiomyocytes (iCMs). To label cells genetically, we used a mouse transgenic line that expresses Cre recombinase under the promoter of the fibroblast-enriched gene, periostin^{15,18,19}. When intercrossed with the R26R-lacZ reporter line²⁰, in which β -galactosidase is activated only in periostin-Cre-expressing cells and their progeny (Fig. 1d–f), we found β -galactosidase activity in many, but not all, cardiac fibroblasts and

¹Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158, USA. ²Department of Pediatrics, University of California, San Francisco, California 94158, USA. ³Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. ⁴Cardiovascular Research Institute, University of California, San Francisco, California 94158, USA. ⁵Department of Medicine, University of California, San Francisco, California 94158, USA. ⁶Developmental Biology and Neonatal Medicine Research Program, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

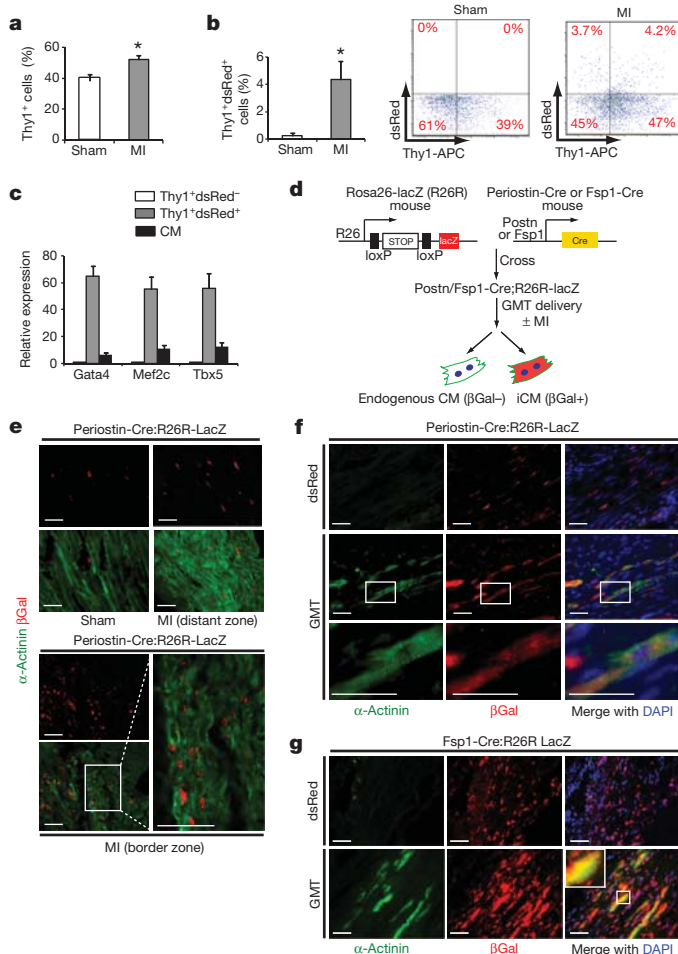


Figure 1 | Genetic lineage tracing demonstrates *in vivo* reprogramming of cardiac fibroblasts to CM-like cells. **a**, Quantification of FACS analyses for Thy1⁺ cells from sham-operated mouse hearts or hearts 2 days after MI ($n = 3$, $*P < 0.05$). **b**, FACS analyses of Thy1⁺ dsRed⁺ cells from sham-operated or post-MI hearts injected with dsRed-expressing retrovirus, with quantification (left) and representative FACS plots (right) ($n = 3$, $*P < 0.05$). APC, allophycocyanin. **c**, qPCR analysis of Gata4, Mef2c and Tbx5 in Thy1⁺ dsRed⁺ cells or endogenous CMs compared to Thy1⁺ dsRed⁻ cells sorted 2 days after post-MI intramyocardial Gata4, Mef2c, Tbx5 and dsRed (GMTR) injection. $n = 3$ with technical quadruplicates. **d**, Schematic diagram showing the genetic fate mapping method to trace the lineage of CMs reprogrammed from periostin-Cre:R26R-lacZ or Fsp1-Cre:R26R-lacZ cells. β Gal, β -galactosidase; Postn, periostin. **e**, Immunofluorescent staining for α -actinin (green) and β -galactosidase (red) on sham-operated or post-MI periostin-Cre:R26R-lacZ mouse hearts 4 weeks post-surgery. Images are from distant or border zones where endogenous CMs were labelled by α -actinin, but were never co-localized with β -galactosidase ($n = 5$ hearts per condition, 8 sections per heart). Scale bar, 50 μ m. **f**, **g**, Immunofluorescent staining for α -actinin, β -galactosidase and DAPI in infarct areas of dsRed- or GMT-injected periostin-Cre:R26R-lacZ (**f**) or Fsp1-Cre:R26R-lacZ (**g**) mouse hearts 4 weeks after MI. Boxed areas indicate regions of magnification. Scale bar, 50 μ m. Error bars indicate standard error of the mean (s.e.m.).

some endocardial and endothelial cells, as previously reported^{15,18,19}. Periostin-Cre activity was absent in bone marrow cells (not shown). β -Galactosidase activity was not detected in any cardiomyocytes, even 4 weeks after injury, consistent with thoracic aortic banding studies, confirming that the periostin-Cre mice mark only descendants of the non-myocyte population, even after MI (Fig. 1e)^{15,18,19}. Isolation of single CMs from these hearts confirmed the absence of β -galactosidase activity in over 1,500 CMs per heart from 6 mice.

In contrast, 4 weeks after MI and retroviral delivery of GMT, numerous β -galactosidase⁺ cells were α -actinin⁺ in the injured areas,

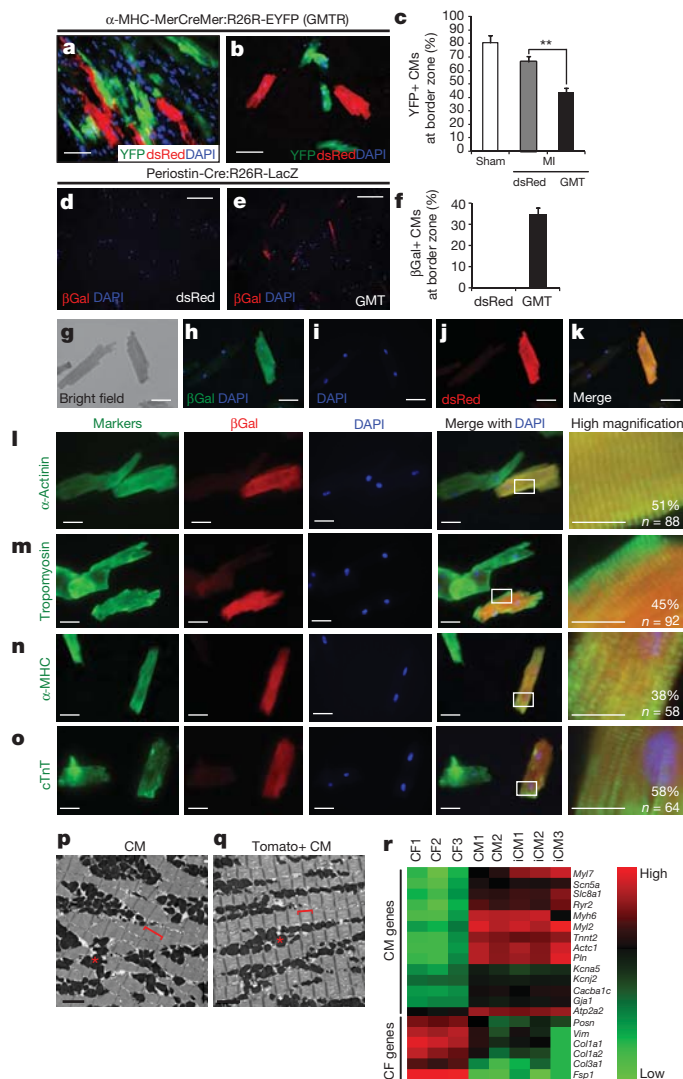
with well-formed sarcomeres and shapes similar to β -galactosidase⁻ myocytes, suggesting that they were descendants of cells that once expressed periostin (Fig. 1f). Similar results were obtained using transgenic mice in which Cre recombinase was under the control of the fibroblast-specific protein 1 (Fsp1; also known as S100a4) promoter²¹ to label the non-myocyte population (Fig. 1g and Supplementary Fig. 2a–f). Thus, analogous to induction of cardiac fibroblasts into skeletal muscle *in vivo* upon introduction of the skeletal muscle master regulator, MyoD²², GMT appeared to induce the formation of cardiomyocytes *in vivo*. We determined if endothelial and circulating haematopoietic cells marked by Tie2-Cre:R26R-lacZ²³ transgenic mice could be reprogrammed to express sarcomeric markers, but found no evidence for such an event (Supplementary Fig. 2g–i).

We formally tested whether retroviral introduction of GMT into non-myocytes could promote cell fusion events in the heart, thereby generating α -actinin⁺ β -galactosidase⁺ cells. We ‘pulse-labelled’ endogenous CMs in transgenic mice with Cre under inducible control of the α -MHC (also known as Myh6) promoter (α -MHC-MerCreMer)²⁴ crossed with R26R-EYFP mice (Supplementary Fig. 3). Subsequently, hearts were injured and infected retrovirally with GMT and dsRed to mark infected dividing cells. After 4 weeks, we detected no YFP⁺ cells co-labelled with dsRed in the GMTdsRed- or dsRed-infected hearts (Fig. 2a, b and Supplementary Fig. 4a). Because pulse labelling marks only ~80% of endogenous CMs in the uninjured heart and ~60% in the infarct border zone²⁵, we quantified the percentage of YFP⁺ pulse-labelled endogenous CMs at the border area. GMT introduction resulted in a reduced percentage of YFP⁺ endogenous CMs compared to total CMs, indicating that the CMs in this region were refreshed by new iCMs (Fig. 2c). These findings suggest that it is unlikely that cell fusion makes a major contribution to the α -actinin⁺ β -galactosidase⁺ cell population, although a minor contribution cannot be ruled out.

If α -actinin⁺ β -galactosidase⁺ cells instead resulted from cellular reprogramming, one might detect progressive stages of reprogramming over time, as cell fusion would yield mature cells soon after fusion without intermediate stages. We therefore analysed heart sections 1, 2, 3 and 4 weeks after injury and GMT infection, and classified cells into four groups based on increasing α -actinin expression and organization into sarcomeres. The number of α -actinin⁺ β -galactosidase⁺ cells in the infarct area increased temporally, as did the maturity of the cells, with progressive increases in the percentage of cells with well-developed sarcomeres (Supplementary Fig. 5).

To avoid false positives from overlaying cells due to the thickness of the heart sections, we isolated adult CMs at the single-cell level from the infarct/border zone of periostin-Cre:R26R-lacZ reprogrammed hearts 4 weeks after coronary ligation (Supplementary Fig. 6a). In this preparation, non-myocytes were removed, and cells were assayed 2–4 h after primary culture. No CMs isolated from dsRed-injected hearts were β -galactosidase⁺ by immunostaining (Fig. 2d). Similarly, CMs from periostin-Cre:R26R-EYFP mice were all YFP⁻, among the thousands of cells visualized, in agreement with the absence of periostin-Cre activity in myocytes after injury. In contrast, 35% of cells in the CM preparation from the border/infarct zone were β -galactosidase⁺ after GMT injection (Fig. 2e, f and Supplementary Fig. 6b). Among the β -galactosidase⁺ cells, 98% were also α -actinin⁺ (Supplementary Fig. 7a–d). Furthermore, in hearts co-injected with GMT and dsRed retrovirus, β -galactosidase⁺ CMs were also positive for dsRed, indicating retroviral infection and their likely origin from non-post-mitotic CMs (Fig. 2g–k).

Most β -galactosidase⁺ cells were large, rod-shaped and binucleated, closely resembling endogenous CMs that were β -galactosidase⁻ from the same preparation. In addition to α -actinin, β -galactosidase⁺ cells expressed multiple sarcomeric markers, including tropomyosin (Fig. 2m), α -MHC (Fig. 2n), and cardiac troponin T (cTnT; also known as Tnnt2) (Fig. 2o). Half of the cells had nearly normal sarcomeric structures throughout the cell. The full spectrum of reprogrammed cells, classified by quality of sarcomeric structure, is shown in



Supplementary Fig. 7. Characterization of single dsRed⁺ YFP[−] cells derived from the α -MHC-Mer-Cre-Mer-YFP pulse-labelled reprogrammed hearts revealed good sarcomere formation and expression of α -actinin, cTnT and connexin 43 (Cx43; also known as Gja1), like dsRed[−] YFP⁺ cells (Supplementary Fig. 4b).

By electron microscopy, about half of the cells from periostin-Cre:R26R-Tomato reprogrammed hearts exhibited well-organized sarcomeres and mitochondria (Fig. 2p, q), although the sarcomeres were consistently shorter than endogenous CMs and their Z-bands more diffuse. Other Tomato⁺ cells displayed sarcomeric organization in parts of the cell and variable mitochondria organization (Supplementary Fig. 8). For simplicity, we will refer to the β -galactosidase⁺ α -actinin⁺ CM-like cells as *in vivo* iCMs, based on morphology and sarcomeric structure.

Finally, we assessed the reprogramming of gene expression in iCMs by qPCR, focusing on the messenger RNA levels of 20 genes normally enriched in mature CMs or cardiac fibroblasts. We tested iCMs isolated from multiple independent hearts alongside cardiac fibroblasts and endogenous CMs. mRNA levels in iCMs were similar to CMs (Fig. 2r and Supplementary Fig. 9), including the downregulation of periostin and *Fsp1*, consistent with the morphological changes described earlier.

In vivo iCMs electrically mature and couple

To determine whether iCMs expressed proteins involved in cell–cell communication similar to endogenous CMs, we examined the

expression pattern of N-cadherin, a cell-surface Ca²⁺-dependent adhesion molecule normally found in intercalated disks within the myocardium²⁶. We found that over 90% of iCMs expressed N-cadherin, with 60% of cells localizing N-cadherin appropriately at the cell border (Fig. 3a). Similarly, about 90% of iCMs expressed Cx43, the major gap junction protein in the heart that promotes electrical coupling and synchronized contraction of myocytes²⁷. Half of the iCMs expressed Cx43 at high levels with good localization relative to endogenous CMs (Fig. 3b), and in 4% of these cells, the Cx43 localization pattern was almost indistinguishable from endogenous CMs (Fig. 3b). Immunohistochemistry also revealed good cell-border localization of Cx43 in iCMs (Fig. 3c).

To assess the function of cell–cell junctions, we imaged the transfer of dyes—microinjected through patch pipettes—between cells, and measured the intercellular transmission of excitation, via Ca²⁺ waves, in small groups of cells isolated from the reprogrammed hearts. Cascade Blue dextran (molecular weight (MW) 10,000), which is too large to pass through gap junctions, was retained in the patched iCMs (identified by the lineage marker periostin-Cre:R26R-Tomato). In contrast, calcein (MW 600) diffused to interconnected endogenous CMs (Fig. 3d). Ca²⁺ waves propagating in an iCM or CM excited intracellular Ca²⁺ release in neighbouring cells (Fig. 3e), suggesting that iCM–CM couplings form functional syncytia (see Supplementary Movie 1).

We next performed recordings from a single-cell suspension of CMs isolated from the border/infarct zone of periostin-Cre:R26R-EYFP mice transduced with GMT, and subsequently compared the action potentials generated by iCMs (YFP⁺) and endogenous CMs (YFP[−]) using standard patch-clamp techniques. Approximately 50% of patched iCMs had a physiological resting membrane potential (−70 mV or less) and 50% of iCMs exhibited contractions in response to electrical stimulation, similar to adult ventricular CMs, which are normally quiescent without stimulation (Fig. 3f and Supplementary Movie 2). Electrophysiology parameters assayed were similar to endogenous ventricular CMs (Fig. 3g). In agreement, intracellular

expression pattern of N-cadherin, a cell-surface Ca²⁺-dependent adhesion molecule normally found in intercalated disks within the myocardium²⁶. We found that over 90% of iCMs expressed N-cadherin, with 60% of cells localizing N-cadherin appropriately at the cell border (Fig. 3a). Similarly, about 90% of iCMs expressed Cx43, the major gap junction protein in the heart that promotes electrical coupling and synchronized contraction of myocytes²⁷. Half of the iCMs expressed Cx43 at high levels with good localization relative to endogenous CMs (Fig. 3b), and in 4% of these cells, the Cx43 localization pattern was almost indistinguishable from endogenous CMs (Fig. 3b). Immunohistochemistry also revealed good cell-border localization of Cx43 in iCMs (Fig. 3c).

To assess the function of cell–cell junctions, we imaged the transfer of dyes—microinjected through patch pipettes—between cells, and measured the intercellular transmission of excitation, via Ca²⁺ waves, in small groups of cells isolated from the reprogrammed hearts. Cascade Blue dextran (molecular weight (MW) 10,000), which is too large to pass through gap junctions, was retained in the patched iCMs (identified by the lineage marker periostin-Cre:R26R-Tomato). In contrast, calcein (MW 600) diffused to interconnected endogenous CMs (Fig. 3d). Ca²⁺ waves propagating in an iCM or CM excited intracellular Ca²⁺ release in neighbouring cells (Fig. 3e), suggesting that iCM–CM couplings form functional syncytia (see Supplementary Movie 1).

We next performed recordings from a single-cell suspension of CMs isolated from the border/infarct zone of periostin-Cre:R26R-EYFP mice transduced with GMT, and subsequently compared the action potentials generated by iCMs (YFP⁺) and endogenous CMs (YFP[−]) using standard patch-clamp techniques. Approximately 50% of patched iCMs had a physiological resting membrane potential (−70 mV or less) and 50% of iCMs exhibited contractions in response to electrical stimulation, similar to adult ventricular CMs, which are normally quiescent without stimulation (Fig. 3f and Supplementary Movie 2). Electrophysiology parameters assayed were similar to endogenous ventricular CMs (Fig. 3g). In agreement, intracellular

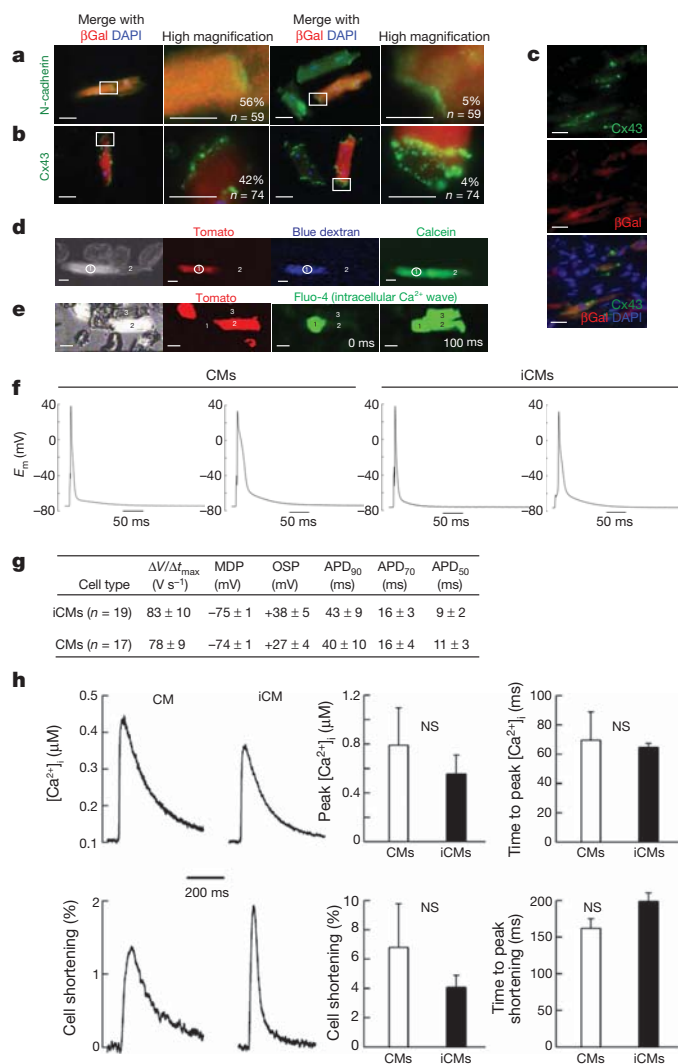


Figure 3 | Electrophysiological properties of iCMs. **a, b**, Immunofluorescent staining for N-cadherin (**a**) or Cx43 (**b**), co-labelled with β -galactosidase (β Gal) and DAPI in isolated CMs from periostin-Cre:R26R-lacZ hearts 4 weeks after injury. Boxed areas are shown in higher magnification with the per cent of cells having the indicated morphology. Green cells represent endogenous CMs, and red/orange cells are iCMs. **c**, Immunohistochemistry for Cx43 on sections from the infarct/border zone of periostin-Cre:R26R-lacZ hearts 4 weeks after GMT injection. Scale bar, 50 μ m in the first and third columns of **a** and **b**, and all of **c**; 20 μ m in the second and fourth columns of **a** and **b**. **d**, Representative images of two CMs in contact with one another, including an iCM (red, cell 1) and an endogenous CM (non-red, cell 2) loaded with large (dextran) or small (calcein) dye. The large blue dextran dye loaded in the iCM (cell 1) by whole-cell patch-clamp method did not travel to the CM (cell 2), but the smaller, gap-junction-permeable dye calcein did cross the cell border ($n = 5$). Scale bar, 50 μ m. **e**, Video frames captured from a group of myocytes, including endogenous CMs (non-red, cells 1 and 3) and an iCM (red, cell 2) imaged for Fluo-4 fluorescence transients corresponding to sarcoplasmic reticulum Ca^{2+} releases. Video frames 100 ms apart show that the Ca^{2+} release has spread throughout the myocyte group, including the iCM ($n = 6$). Scale bar, 50 μ m. **f**, Intracellular electrical recording of *in vivo*-derived YFP⁺ iCMs and endogenous YFP⁻ CMs from the same preparation. E_m , membrane potential in millivolts. **g**, Table of action potential parameters measured for CMs and iCMs, including maximum upstroke velocity ($\Delta V/\Delta t_{\max}$) and minimum diastolic potential (MDP) measured immediately preceding stimulation, overshoot potential (OSP), and the APDs at 90, 70 and 50% repolarization. **h**, Characteristic single-field-stimulated intracellular calcium transients ($[\text{Ca}^{2+}]_i$) recorded from endogenous (left) or induced (right) CMs. Bottom, the simultaneously recorded per cent cell shortening responses triggered by the Ca^{2+} transients, in the same two cells. Quantifications from 6 iCMs and 4 endogenous CMs are shown in the right four panels. NS, not significant. For experiments performed in **d–h**, cells were isolated from periostin-Cre:Rosa-YFP mice 8 weeks post-MI and virus transduction. Error bars indicate s.e.m.

calcium releases and cell shortening in iCMs were comparable to endogenous CMs (Fig. 3h). The distribution of action potential durations (APDs) was bimodal in iCMs and CMs, suggesting that reprogrammed cells were incorporated near the epicardial (short APD₉₀ (action potential duration at 90% of repolarization)) and endocardial (long APD₉₀) sides of ventricular tissue (Supplementary Fig. 10).

In vivo GMT improves cardiac function

Because *in vivo* reprogrammed iCMs had contractile potential and electrically coupled with viable endogenous CMs (and other iCMs), we asked whether converting endogenous non-myocytes into new myocytes translates into partial restoration of heart function after MI. All studies were performed in a blinded fashion, including the retroviral injections, and were decoded only after completion of the measurements. By Evans blue/triphenyltetrazolium chloride (TTC) double staining, the area at risk (AAR) and the infarct size were similar in GMT- or dsRed-injected mice 48 h after coronary ligation (Supplementary Fig. 11a). Three months after MI, cardiac function was examined by magnetic resonance imaging (MRI). The fraction of blood ejected with each ventricular contraction (ejection fraction), the volume of blood ejected (stroke volume), and the total cardiac output per minute were significantly improved in GMT-infected mice, particularly the stroke volume and cardiac output, possibly due to cardiac enlargement (Fig. 4a). To determine the time course of these improvements, other mice underwent serial high-resolution two-dimensional echocardiography 1 day before MI, and 3 days, 1, 4, 8 and 12 weeks after MI (Supplementary Fig. 11b–d). All mice showed a comparable

reduction in left ventricular function after coronary artery ligation (Supplementary Fig. 11c). Although different imaging approaches yield different absolute value norms, the overarching trends observed by echocardiography were similar to our MRI findings, in that functional improvements for all parameters were statistically significant 8 and 12 weeks after injection (Supplementary Fig. 11c).

We next performed qPCR to monitor the expression levels of atrial natriuretic factor, brain natriuretic peptide and tenascin C in injured and control hearts. We found that MI led to the upregulation of all three peptides, but this upregulation was attenuated in GMT-injected infarcted hearts (Fig. 4b). Expression levels of collagen genes, which were increased in dsRed-infected MI hearts, were also partially restored by injecting GMT (Fig. 4c). Furthermore, the scar area calculated from 16 sections at four levels of the heart was significantly smaller 8 weeks after MI in the GMT-treated group. To determine if the muscle cells in the scar area were reprogrammed iCMs, we repeated the experiments in periostin-Cre:R26R-LacZ transgenic mice. α -Actinin⁺ cells in the scar area were also β -galactosidase positive, suggesting that they were newly born iCMs of non-myocyte origin (Fig. 4e). Vascular density was significantly increased in the border zone of reprogrammed hearts at 8 weeks (Supplementary Fig. 12). Electrocardiographic (ECG) studies (telemetry) over a 24-h period did not indicate evidence for more arrhythmias in GMT- versus dsRed-injected control mice, and no mice suffered sudden death (not shown).

Thymosin β 4 enhances effects of GMT *in vivo*

We hypothesized that infecting more Thyl⁺ cells would enhance functional improvement. Thymosin β 4, a 43-amino-acid G-actin monomer-binding protein, promotes cell migration^{28,29}, cardiac cell survival^{28,30} and activates epicardial cells to become more proliferative and yield more cardiac fibroblasts and endothelial cells^{31,32}. It also improves cardiac function and decreases scar size after MI²⁸. To test cardiac fibroblast migration, we used a cardiac explant migration assay^{7,28}. The average time for fibroblasts to migrate from adult heart explants was 3 weeks; however, thymosin β 4 treatment led to equivalent fibroblast migration within 2 weeks and within only 3 days in

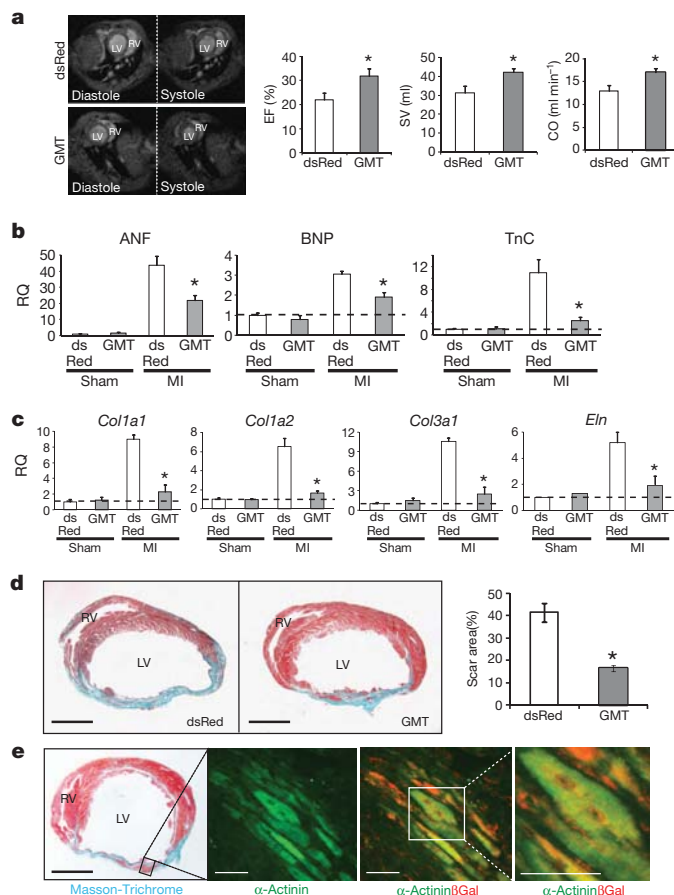


Figure 4 | *In vivo* delivery of cardiac reprogramming factors improves cardiac function after myocardial infarction. **a**, Ejection fraction (EF), stroke volume (SV) and cardiac output (CO) of the left ventricle were quantified by MRI 12 weeks after MI ($n = 9$ for each group, $*P < 0.05$). Left four panels show representative transverse images of the thorax, containing hearts at the end of diastole (relaxation) or systole (contraction) from dsRed- or GMT-injected mice, compared to sham-operated age- and strain-matched controls. LV, left ventricle; RV, right ventricle. **b**, qPCR of atrial natriuretic factor (ANF), brain natriuretic peptide (BNP) and tenascin (TnT) on RNA extracted from the border zone of hearts 4 weeks after MI and injection of dsRed or GMT. RQ, relative quantification. **c**, qPCR of collagen type I alpha 1 (*Col1a1*), *Col1a2*, *Col3a1* and elastin (*Eln*) on RNA extracted from the border zone of hearts 4 weeks after MI and injection of dsRed or GMT. Data in **b** and **c** are relative to dsRed-injected sham-operated mice, indicated by the dashed line. $n = 3$ for each genotype with technical quadruplicates. $*P < 0.05$. **d**, Masson-Trichrome staining on heart sections 8 weeks post-MI injected with dsRed or GMT with quantification of scar size. Scale bars, 500 μ m. dsRed, $n = 8$; GMT, $n = 9$; $*P < 0.05$. **e**, Masson-Trichrome (left) and immunofluorescent staining for α -actinin and/or β -galactosidase (β Gal; right) in GMT injected periostin-Cre:R26R-lacZ mouse heart 4 weeks post-surgery. Scale bars, 500 μ m in the left panel, 50 μ m in the right three panels. Error bars indicate s.e.m.

heart explants taken after MI (Supplementary Fig. 13a). Similarly, vimentin⁺ cell proliferation was even more pronounced post-MI in the presence of thymosin β 4, as marked by phosphohistone H3 (Supplementary Fig. 13b). The percentage of Thyl⁺ (Supplementary Fig. 13c) or vimentin⁺ (Supplementary Fig. 14a) cells infected by retrovirus after MI doubled upon intramyocardial thymosin β 4 injection ($316,607 \pm 17,799$ dsRed⁺ cells averaged from three hearts). GMT did not increase the number of cardiac fibroblasts (Supplementary Fig. 13d). Delivering GMT-expressing retrovirus to more cells by adding thymosin β 4 yielded more β -galactosidase⁺ iCMs compared to total CMs in single-cell CM culture from the infarct/border zone of periostin-Cre:R26R-lacZ hearts (51% versus 35%) (Supplementary Fig. 13e). However, we observed no change in the degree or efficiency of *in vivo* reprogramming, with the percentage of iCMs generated from

the total GMT-infected cell population remaining at $\sim 12\%$ (Supplementary Figs 13e and 14b).

Injecting thymosin β 4 immediately after ligation improved cardiac function, as previously reported^{28,30}. Co-injecting thymosin β 4 and GMT further improved ejection fraction and cardiac output 8 weeks after infarction (Supplementary Figs 13f and 14d, e). Furthermore, co-injecting thymosin β 4 and GMT caused less scarring than injecting either alone (Supplementary Fig. 13g), despite similar areas at risk and initial infarct sizes (Supplementary Fig. 14c).

Discussion

We show that upon cardiac injury, resident cardiac non-myocytes—primarily fibroblasts—can be converted into CM-like cells *in vivo* following local delivery of GMT by retroviral-mediated gene transfer. *In vivo* cardiac reprogramming occurred with similar initial efficiency as observed *in vitro* (10–15%). However, *in vivo* iCMs were more fully reprogrammed and more closely resembled endogenous CMs than their cultured counterparts. This may result from factors within the native microenvironment—including extracellular matrix, secreted proteins, and tissue stiffness—that further enhance reprogramming. Improved cardiac function may be explained by the diversion of a small percentage of fibroblasts into new CM-like cells, suggesting functional integration of these muscle cells. Although non-myocytes convert to iCMs to help regenerate the damaged heart, alteration of fibroblast behaviour by GMT may contribute to the effects on scar formation and cardiac function. Although it is difficult to separate the relative contributions of new muscle formation and other non-cell-autonomous effects, non-myocyte reprogramming in the heart appears to be beneficial for cardiac function. Optimizing gene delivery to more cells would probably enhance functional benefits.

Improvement upon thymosin β 4 addition is in agreement with the notion that increasing the delivery of GMT to more cells could enhance cardiac repair. Pre-treating hearts with thymosin β 4 several days before injury resulted in a small population of epicardial-derived cells that could behave as myocyte precursors, but not if thymosin β 4 was given at the time of injury³³. Nevertheless, transduction of GMT into these progenitors, or other rare progenitors yet to be identified, might promote their differentiation into cardiomyocytes. Because thymosin β 4 is also pro-angiogenic^{31,32}, the cooperativity between GMT and thymosin β 4 may be multifaceted and will be interesting to explore.

The ability to regenerate adult heart tissue from endogenous cells is a promising approach to treating cardiac disease that may face fewer obstacles to clinical translation than other approaches. Improving the delivery of reprogramming factors, using small molecules and epigenetic modulators, and conducting trials in large animals will be important to refine the technology and assess its safety and efficacy, particularly regarding arrhythmias.

METHODS SUMMARY

Retroviruses. Retroviruses were generated as described⁷ using pMXs retroviral vectors containing coding regions of Gata4, Mef2c, Tbx5 and dsRed. Ultra-high titre virus ($>1 \times 10^{10}$ plaque-forming units (p.f.u.) per ml) was obtained by ultracentrifugation.

Animals, surgery, echocardiography and electrocardiography. Periostin-Cre:R26R-lacZ or Fsp1-Cre:R26R-lacZ mice were obtained by crossing periostin-Cre mice¹⁹ or Fsp1-Cre mice²¹ and R26R-lacZ mice²⁰. Periostin-Cre:R26R-EYFP or α -MHC-MerCreMer:R26R-EYFP mice were obtained by crossing periostin-Cre or α -MHC-MerCreMer²⁴ mice and R26R-EYFP mice. Periostin-Cre:R26R-Tomato mice were obtained by crossing periostin-Cre mice and R26R-Tomato mice. Surgeries and subsequent analyses were performed blinded to genotype and intervention. MI was induced by permanent ligation of the left anterior descending artery (LAD) as described³⁴. A pool of concentrated virus (GMT or GMT-R) was mixed, and 10 μ l of mixed virus plus 10 μ l of PBS or 40 ng μ l⁻¹ thymosin β 4 were injected along the boundary between the infarct and border zones. Mouse echocardiography and surface electrocardiography were performed as described³⁴. All mouse work was done with the approval of the University of California, San Francisco (UCSF) animal care oversight committee.

Immunohistochemistry, immunocytochemistry and electron microscopy. Immunohistochemistry, immunocytochemistry and electron microscopy were performed as described^{34,35}. Scar size was determined by Masson-Trichrome staining^{28,34}. The AAR and myocardial infarct size were determined by Evans blue/TTC labelling³⁴.

Cardiomyocyte isolation, patch-clamp, and fibroblast migration assays. Adult cardiomyocytes were isolated as described with minor modifications³⁶. Single-cell patch-clamp recordings were performed as described³⁷. Migration assays were performed according to published protocols^{7,28}.

FACS and quantitative RT-PCR. Dissociated cardiac cells were stained with APC-conjugated anti-Thy1 antibody (eBioscience). Stained cells were sorted by FACSARIA2 (BD) and RNA extracted in TRIzol (Invitrogen). qPCR was performed using ABI 7900HT (TaqMan, Applied Biosystems).

Statistics. Differences between groups were examined for statistical significance using unpaired Student's *t*-tests or ANOVA. *P* < 0.05 was regarded as significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 March 2011; accepted 14 March 2012.

Published online 18 April 2012.

- Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661–680 (2008).
- Passier, R., van Laake, L. W. & Mummery, C. L. Stem-cell-based therapy and lessons from the heart. *Nature* **453**, 322–329 (2008).
- Srivastava, D. & Ivey, K. N. Potential of stem-cell-based therapies for heart disease. *Nature* **441**, 1097–1099 (2006).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
- Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
- Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to β -cells. *Nature* **455**, 627–632 (2008).
- Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–389 (2011).
- Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
- Ieda, M. *et al.* Cardiac fibroblasts regulate myocardial proliferation through β 1 integrin signaling. *Dev. Cell* **16**, 233–244 (2009).
- Byun, J. *et al.* Myocardial injury-induced fibroblast proliferation facilitates retroviral-mediated gene transfer to the rat heart *in vivo*. *J. Gene Med.* **2**, 2–10 (2000).
- Snider, P. *et al.* Origin of cardiac fibroblasts and the role of periostin. *Circ. Res.* **105**, 934–947 (2009).
- Baudino, T. A., Carver, W., Giles, W. & Borg, T. K. Cardiac fibroblasts: friend or foe? *Am. J. Physiol. Heart Circ. Physiol.* **291**, H1015–H1026 (2006).
- Camelliti, P., Borg, T. K. & Kohl, P. Structural and functional characterisation of cardiac fibroblasts. *Cardiovasc. Res.* **65**, 40–51 (2005).
- Snider, P. *et al.* Periostin is required for maturation and extracellular matrix stabilization of noncardiomyocyte lineages of the heart. *Circ. Res.* **102**, 752–760 (2008).
- Takeda, N. *et al.* Cardiac fibroblasts are essential for the adaptive response of the murine heart to pressure overload. *J. Clin. Invest.* **120**, 254–265 (2010).
- Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
- Bhowmick, N. A. *et al.* TGF- β signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* **303**, 848–851 (2004).
- Murry, C. E., Kay, M. A., Bartosek, T., Hauschka, S. D. & Schwartz, S. M. Muscle differentiation during repair of myocardial necrosis in rats via gene transfer with MyoD. *J. Clin. Invest.* **98**, 2209–2217 (1996).
- Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: A new model for endothelial cell-lineage analysis *in vivo*. *Dev. Biol.* **230**, 230–242 (2001).
- Sohal, D. S. *et al.* Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circ. Res.* **89**, 20–25 (2001).
- Hsieh, P. C. *et al.* Evidence from a genetic fate-mapping study that stem cells refresh adult mammalian cardiomyocytes after injury. *Nature Med.* **13**, 970–974 (2007).
- Li, J. *et al.* Cardiac-specific loss of N-cadherin leads to alteration in connexins with conduction slowing and arrhythmogenesis. *Circ. Res.* **97**, 474–481 (2005).
- Saffitz, J. E., Laing, J. G. & Yamada, K. A. Connexin expression and turnover: implications for cardiac excitability. *Circ. Res.* **86**, 723–728 (2000).
- Bock-Marquette, I., Saxena, A., White, M. D., Dimaio, J. M. & Srivastava, D. Thymosin β 4 activates integrin-linked kinase and promotes cardiac cell migration, survival and cardiac repair. *Nature* **432**, 466–472 (2004).
- Goldstein, A. L., Hannappel, E. & Kleinman, H. K. Thymosin β 4: actin-sequestering protein moonlights to repair injured tissues. *Trends Mol. Med.* **11**, 421–429 (2005).
- Hinkel, R. *et al.* Thymosin β 4 is an essential paracrine factor of embryonic endothelial progenitor cell-mediated cardioprotection. *Circulation* **117**, 2232–2240 (2008).
- Smart, N. *et al.* Thymosin β 4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* **445**, 177–182 (2007).
- Bock-Marquette, I. *et al.* Thymosin β 4 mediated PKC activation is essential to initiate the embryonic coronary developmental program and epicardial progenitor cell activation in adult mice *in vivo*. *J. Mol. Cell. Cardiol.* **46**, 728–738 (2009).
- Smart, N. *et al.* *De novo* cardiomyocytes from within the activated adult heart after injury. *Nature* **474**, 640–644 (2011).
- Qian, L. *et al.* miR-24 inhibits apoptosis and represses Bim in mouse cardiomyocytes. *J. Exp. Med.* **208**, 549 (2011).
- Larsen, T. H., Saetersdal, T. & Grong, K. The ultrastructure of the myocyte in different regions of experimental infarcts in the cat heart. *Res. Exp. Med. (Berl.)* **186**, 295–306 (1986).
- Xu, H., Guo, W. & Nerbonne, J. M. Four kinetically distinct depolarization-activated K^+ currents in adult mouse ventricular myocytes. *J. Gen. Physiol.* **113**, 661–678 (1999).
- Fenske, S. *et al.* HCN3 contributes to the ventricular action potential waveform in the murine heart. *Circ. Res.* **109**, 1015–1023 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for technical assistance from the Gladstone Histology Core (C. Miller), Gladstone Genomics Core (L. Ta, Y. Hao, B. Chadwick), UCSF MRI Core (M. Wendland, J. Hawkins) and Laboratory for Cell Analysis at UCSF (S. Elmes). We thank all the members of the Srivastava laboratory for helpful discussions; G. Howard and B. Taylor for editorial help; and B. Bruneau and B. Conklin for helpful discussions and critical reviews of the manuscript. We also thank J. Nerbonne, N. Foeger, and members of the Nerbonne laboratory for assistance with the adult myocyte isolation protocol. L.Q. is a postdoctoral scholar of the California Institute for Regenerative Medicine (CIRM). V.V. is supported by grants from the GlaxoSmithKline Research and Education Foundation and the NIH/NHLBI (K08HL101989). J.-d.F. is supported by a postdoctoral fellowship from American Heart Association. S.J.C. was supported by R01 HL060714 from NHLBI/NIH. D.S. was supported by grants from NHLBI/NIH, CIRM, the Younger Family Foundation, Roddenberry Foundation and the L.K. Whittier Foundation. This work was supported by NIH/NCRR grant (C06 RR018928) to the Gladstone Institutes.

Author Contributions L.Q. designed, supervised and performed the experiments. Y.H. performed all surgeries, echoes and ECGs, and contributed to tissue sectioning and sample preparation. C.I.S. performed all cellular electrophysiology experiments. A.F. quantified scar size and induced CMs and helped with mouse colony maintenance. V.V. helped with isolation of adult CMs and implantation of transmitters. S.J.C. provided periostin-Cre:Rosa26-lacZ mice and supplemental data. J.-d.F. provided initial reagents and technical knowledge and helpful discussion. D.S. designed and supervised the work. L.Q. and D.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.S. (dsrivastava@gladstone.ucsf.edu).

METHODS

Retrovirus generation, concentration, and titration. Retroviruses were generated as described⁷. To generate virus, pMXs retroviral vectors containing the coding regions of Gata4, Mef2c, Tbx5 and dsRed were transfected into Plat-E cells using Eugene 6 (Roche). Forty-eight hours after transfection, virus-containing supernatants were collected and concentrated by standard ultracentrifugation. Retroviral titration was performed using the Retro-X qRT-PCR Titration Kit (Clontech), as per the manufacturer's protocols. Ultra-high titre virus ($>1 \times 10^{10}$ plaque-forming units (p.f.u.) per ml) was resuspended in PBS. After verification of high transduction efficiency in cell culture ($>95\%$), a large number of small stock aliquots (10 μ l) were made and frozen at -80°C to ensure consistency among experiments. After one freeze-thaw cycle, titrations were repeated to ensure that active virus was maintained at the desired 1×10^{10} p.f.u. concentration for *in vivo* injection.

Mouse lines. Periostin (Postn)-Cre:R26R-lacZ mice were obtained by crossing Postn-Cre mice and Rosa26-lacZ mice. Postn-Cre:R26R-YFP mice were obtained by crossing Postn-Cre mice and Rosa26-EYFP mice, and Postn-Cre:R26R-Tomato mice were obtained by crossing Postn-Cre mice and Rosa26-Tomato mice. All transgenic lines for immunohistochemistry and single cell isolation were maintained by crossing with C57BL6 mice (Charles River). BALB/C mice (Charles River) were used for all functional studies after permanent ligation of the left anterior descending artery (LAD) and virus injection. Fsp1-Cre, Tie2-Cre and Myh6-MerCreMer mice were obtained from Jackson Labs, and lines were validated before further breeding. Fsp1-R26R, Tie2-R26R and Myh6-MerCreMer-YFP mice were obtained by crossing Fsp1-Cre, Tie2-Cre or Myh6-MerCreMer mice to R26R-lacZ or R26R-EYFP mice. Efficiency of Cre recombination induction for Myh6-MerCreMerYFP was tested by immunohistochemistry for YFP after injection of various doses of tamoxifen. To pulse label the pre-existing CMs, adult Myh6-MerCreMerYFP mice (8–12-weeks old) were treated with tamoxifen (Sigma) by intraperitoneal injection once a day for 5 days at a dosage of $20 \text{ mg kg}^{-1} \text{ day}^{-1}$. GMT delivery and coronary artery ligation were performed 2 days afterwards.

Mouse MI model and *in vivo* delivery. The animal protocol for surgery was approved by institutional guidelines (UCSF Institutional Animal Care and Use Committee). All surgeries and subsequent analyses were performed blinded for genotype and intervention. Mice were anaesthetized with 2.4% isoflurane/97.6% oxygen and placed in a supine position on a heating pad (37°C). Animals were intubated with a 19 G stump needle and ventilated with room air using a MiniVent Type 845 mouse ventilator (Hugo Sachs Elektronik-Harvard Apparatus; stroke volume, 250 μ l; respiratory rate, 120 breaths per minute). MI was induced by permanent ligation of the LAD with a 7-0 prolene suture as described²⁸. Sham-operated animals served as surgical controls and were subjected to the same procedures as the experimental animals with the exception that the LAD was not ligated. A pool of concentrated virus (GMT, or GMTR) was mixed, and 10 μ l of mixed virus plus 10 μ l of PBS or 40 ng μ l⁻¹ thymosin β 4 was injected into the myocardium through an insulin syringe with an incorporated 29 G needle (BD). Injection with a full dosage was carried out along the boundary between the infarct zone and border zone based on the blanched infarct area after coronary artery occlusion. After injection, the chest was closed with sutures and the mouse was allowed to recover with the mouse ventilator and heating pad. All surgical procedures were performed under aseptic conditions. At 2 days and 1, 2, 4, 8 and 12 weeks after occlusion and viral delivery, the hearts were removed for perfusion fix in 4% paraformaldehyde (PFA) for preparation of paraffin sections for structural analysis and immunohistochemistry or in 0.5% PFA in 5% sucrose followed by cryostat sectioning for immunofluorescent staining. Concurrently, heart tissues within the infarct zone, border zone and non-ischaemic zone distal to the infarct zone were dissected for RNA or protein isolation.

Determination of the AAR and MI size. At 48 h after coronary ligation, the mice were anaesthetized and cannulated with tubing. Evans blue (2%, Sigma) was perfused into the aorta, thus all myocardial tissue was stained blue except the AAR. The left ventricle was isolated and cut into four $\sim 1 \text{ mm}$ pieces with the first cut at the ligation level. Left ventricle slices were stained in 1.5% TTC for 30 min at 37°C , and then fixed in 4% PFA overnight at 4°C . The area of infarction was demarcated as a white area, whereas viable myocardium was stained red. Photographs were taken from both sides of each section. The AAR and the infarct area were determined via planimetry with the computer software ImagePro (Biorad). Infarct size was calculated as the percentage of MI compared with the AAR using the described methodology²⁸.

Determination of scar size. Standard Masson-Trichrome staining was performed on hearts 8 weeks post-viral delivery and coronary artery ligation. To determine the scar size, we used ImagePro software to measure the scar area (blue) and healthy area (red) on transverse sections spanning four levels (50 mm between two levels, with the first level starting right below the ligation) within the left ventricle of a MI heart. From each level, we measured four slices of

tissues as technical quadruplicates (for a total of 16 sections). The averaged number was used for statistics and comparison. The measurements and calculations were conducted in a blinded manner.

Mouse echocardiography. Echocardiography was performed by the Vevo 770 High-Resolution Micro-Imaging System (VisualSonics) with a 15-MHz linear array ultrasound transducer. The left ventricle was assessed in both parasternal long-axis and short-axis views at a frame rate of 120 Hz. End-systole or end-diastole was defined as the phase in which the smallest or largest area of the left ventricle, respectively, was obtained and used for ejection fraction measurement. Left ventricular end-systolic diameter and left ventricular end-diastolic diameter were measured from the left ventricular M-mode tracing with a sweep speed of 50 mm s^{-1} at the papillary muscle level for calculating the shortening fraction. B-mode was used for two-dimensional measurements of end-systolic and end-diastolic dimensions.

Mouse surface electrocardiography. Mice were anaesthetized with 1.75% isoflurane at a core temperature of $37\text{--}38^\circ\text{C}$. Four needle electrodes (AD Instruments) were placed subcutaneously in standard limb lead configurations. For each mouse, 10–20 s of continuous signals were sampled at 10 kHz in each lead configuration with a PowerLab4/30 interface (AD Instruments). Data analysis was performed offline with electronic calipers on averaged beats (Chart5Pro v5.4.2, AD Instruments).

Mouse awake electrocardiography. To record awake electrocardiograms in six reprogrammed post-MI mice and six control post-MI mice, transmitters were surgically implanted according to the manufacturer's instructions (Data Sciences International). After a 3-day recovery period, the electrocardiogram was recorded continuously for 48 h in each mouse. Tracings were analysed off-line and were scored by a blinded investigator for the presence and frequency of arrhythmias.

MRI. MRI was performed on a Varian DirectDrive 7T small-animal scanner. Each mouse was anaesthetized by inhalation of 2% isoflurane/98% oxygen administered via an MR-compatible mobile inhalation anaesthesia system (Vet Equip). The mice were put in supine position on a homemade heating bed to keep the temperature at 37°C . Two ECG leads were inserted into the right front and left rear leg. ECG waveforms were monitored with a small animal monitoring and gating system (SA instruments). The mouse was then placed into a homemade 1H birdcage coil with an inner diameter of 32 mm. A group of ECG- (R-wave rising edge) triggered spin echo scout images were acquired first to define the oblique plane of the short axis. Then an ECG-triggered two-dimensional gradient echo sequence with an echo time of 2.75 ms, repetition time of 200 ms and a flip angle of 45° was used to obtain nine short-axis images at 12 or 13 phases per cardiac cycle. Each scan consisted of 8–9 contiguous slices spanning the left ventricle from apex to base with 1-mm thickness, a matrix size of 128×128 , a field of view of $25.6 \times 25.6 \text{ mm}$, and four averages.

Isolation of adult CMs. Adult CM isolation was performed as described with minor modifications³⁶. Briefly, adult mice were anaesthetized with isoflurane and mechanically ventilated. Hearts were removed and perfused retrogradely via aortic cannulation with a constant flow of 3 ml min^{-1} in a Langendorff apparatus. Hearts were perfused at 37°C for 5 min with supplemented Wittenberg Isolation Medium (WIM) containing (in mM): 116 NaCl, 5.4 KCl, 6.7 MgCl_2 , 12 glucose, 2 glutamine, 3.5 NaHCO_3 , 1.5 KH_2PO_4 , 1.0 NaH_2PO_4 , 21 HEPES, with 1.5 nM insulin, essential vitamins (GIBCO), and essential amino acids (GIBCO) (pH 7.4), followed by digestion solution (WIM, supplemented with 0.8 mg ml^{-1} collagenase II and $10 \mu\text{M}$ CaCl_2) for 10 min (4 min for paired CMs that were used in cell-cell coupling experiments). Hearts were then removed from the Langendorff apparatus while intact (with tissues loosely connected). Desired areas (that is, border/infarct zone) were then micro-dissected under the microscope, followed by mechanical dissociation, triturating, and resuspension in a low-calcium solution (WIM, supplemented with 5 mg ml^{-1} BSA, 10 mM taurine, and $150 \mu\text{M}$ CaCl_2). Cells were then spun at low speed, supernatant was removed, and calcium was gradually reintroduced through a series of washes. For electrophysiology experiments, cells were used on the same day as isolation and until recording were stored at room temperature (21°C) in M199 (Gibco) supplemented with 5 mM creatine, 2 mM L-carnitine, 5 mM taurine and 1.5 nM insulin. For immunohistochemistry, cells were plated onto laminin-coated culture slides, allowed to adhere, and fixed on the day of isolation. For electron microscopy or qPCR, iCMs were selected manually by micro-pipette based on the presence of periostin-Cre:R26R-YFP/Tomato signal under the fluorescent microscope right after isolation.

Cardiac fibroblast migration assay. The migration assay was performed according to the explant culture protocol as described¹³. In brief, isolated adult mouse hearts were minced into small pieces less than 1 mm^3 in size. The explants were plated on gelatin-coated dishes and cultured in explant medium (IMDM, 20% FBS) until fibroblasts migrated out from minced tissue. The time required for ten heart pieces to have migratory fibroblasts surrounding them were recorded.

FACS analyses and sorting. At 48 h after LAD and viral introduction, hearts were removed and minced into small pieces less than 1 mm^3 in size. Blood cells and debris were removed by several washings of PBS. Minced cardiac tissues were digested in an eppendorf tube and shaken with glass beads in enzyme buffer (collagenase/dispase plus DNaseI; Roche) at 37°C . After passing through a $40\text{-}\mu\text{m}$ cell strainer, dissociated cardiac cells were stained with APC-conjugated anti-Thy1 antibody (eBioscience) for 30 min at room temperature. After washing with PBS twice, stained cells were sorted by FACSaria2 (BD).

Immunohistochemistry. For immunofluorescence, after perfusion-fixed hearts were taken out and fixed in 0.5% PFA overnight, ventricles below the ligation were embedded in OCT compound and frozen in liquid nitrogen. Sections were blocked in Universal Blocking Buffer (BioGenex) for 10 min, and then stained with primary antibodies against α -actinin (Sigma Aldrich), vimentin (Progen), β -galactosidase (Abcam), pH3 (Millipore), RFP (Biovision), Thy1 (BD), CD34 (Abcam), WT1 (Abcam), PECAM (BD), NG2 (Millipore) and GFP (Invitrogen) for 1 h at room temperature. After washing three times with PBST (PBS plus 0.1% Triton), sections were incubated in secondary antibodies for 1 h followed by washing an additional three times with PBST. Finally, the sections were mounted in Vectashield with DAPI (Vector Laboratories).

Immunocytochemistry. Isolated cells plated on chamber slides were fixed in 4% PFA at 4°C overnight and washed with PBS twice. Cells were then incubated with primary antibodies against sarcomeric α -actinin (Sigma Aldrich), cTnT (Thermo Scientific), RFP (Biovision), tropomyosin (Hybridoma Bank), MF20 (Hybridoma Bank), Cx43 (Sigma Aldrich), N-cadherin (Invitrogen), β -galactosidase (Abcam), smooth muscle actin (Sigma Aldrich) and vimentin (Progen) for 1 h at room temperature and washed with PBS three times, then incubated with secondary IgG antibodies conjugated to Cy 488 or 594 (Jackson ImmunoResearch) for 30 min. After washing with PBS, cells were mounted in Vectashield with DAPI.

Electron microscopy. For electron microscopy, cells were fixed in 2% glutaraldehyde, 1% paraformaldehyde in 0.1 M sodium cacodylate buffer, pH 7.4, post fixed in 2% osmium tetroxide in the same buffer, *en bloc* stained with 2% aqueous uranyl acetate, dehydrated in acetone, infiltrated, and embedded in LX-112 resin (Ladd Research Industries). Samples were ultrathin sectioned on a Reichert Ultracut S ultramicrotome and counter stained with 0.8% lead citrate. Grids were examined on a JEOL JEM-1230 transmission electron microscope (JEOL USA) and photographed with the Gatan Ultrascan 1000 digital camera (Gatan).

Action potential recordings. Isolated myocytes suspended in tissue culture medium were transferred to a superfusion chamber (RC-26GLP; Warner Instruments) on the stage of a Nikon TiS inverted fluorescence microscope equipped with a dual wavelength microfluorometer (IonOptix). Myocytes isolated from either periostin-Cre:R26R-YFP or periostin-Cre:R26R-Tomato hearts after GMT infection were identified as control CMs or iCMs on the basis of Tomato fluorescence (Texas red optics with exciter $560 \pm 55\text{ nm}$, emitter $645 \pm 75\text{ nm}$) or YFP (standard FITC optics), and were chosen for study if they lacked spontaneous beating and partial contractions. The myocytes were whole-cell patch-clamped using an Axopatch 200B amplifier and pClamp software (Molecular Devices). Patch electrodes of 2–5 M Ω (1B-150F; WPI) were filled with intracellular solution containing 120 mM KCl, 20 mM NaHEPES, 10 mM MgATP, 5 mM K₂EGTA (or 0.1 mM), 2 mM MgCl₂, and adjusted to pH 7.1 with KOH. The cells were superfused with a modified Tyrode's extracellular solution containing 137 mM NaCl, 10 mM NaHEPES, 10 mM dextrose, 5 mM KCl, 2 mM CaCl₂, 1 mM MgCl₂, adjusted to pH 7.4 with NaOH. After G Ω seal formation, whole-cell access to the myocyte was established by applying brief pressure pulses, and the amplifier was switched to current clamp mode, whereupon the cell's resting potential developed. Action potentials were stimulated at 0.33 Hz using 2 nA, 2 ms current pulses applied through the patch pipette, and were signal averaged in tens. All membrane potentials were corrected for a -5.6 mV liquid junction potential determined via pClamp software. Finally, the amplifier was switched back to voltage clamp mode to identify individual ion channel currents as required. In some experiments, the concentration of K₂EGTA in the intracellular solution was reduced to 0.1 mM to permit excitation-contraction coupling to occur, and 100 μM K5Fluo-4 (Invitrogen Corp) was added to define the presence of cytosolic Ca²⁺ transients during action potentials. Electrophysiology data were digitized at 5 kHz and low-pass filtered at 2 kHz. Analysis was performed using pClamp, Microsoft Excel, and Origin (OriginLab) software. Action potential duration was measured from the point of maximum depolarizing voltage change ($\Delta V/\Delta t_{\text{max}}$) to 50, 70 and 90% repolarization. Unless stated, experiments were performed at room temperature.

Field stimulation experiments. Isolated myocytes were loaded with Fluo-4 for 30 min at room temperature before being transferred to the superfusion chamber. The loading solution contained a 1:10 mixture of 5 mM Fluo-4 AM in dry DMSO and PowerloadTM concentrate (Invitrogen) which was diluted 100-fold into extracellular Tyrode's solution containing suspended myocytes. An additional 20 min was allowed for de-esterification before commencing recordings. Contractions and Ca²⁺ transients were evoked by applying voltage pulses at 0.33 Hz, between platinum wires placed on either side of the cell of interest and connected to a field stimulator (IonOptix, Myopacer). The pulses were of 2 ms duration and set at 150% of the threshold required to elicit twitches. Fluo-4 fluorescence transients were recorded via a standard filter set (#49011 ET, Chroma Technology) in batches of ten to enable signal averaging. Between stimuli, the fluorescence excitation light was blocked by an electromechanical shutter (CS35; Vincent Associates). Resting fluorescence was recorded after cessation of pacing, and background light was obtained after picking up and removing the cell from the field of view with a patch electrode at the end of the experiment. The Ca²⁺ transients were calibrated using the pseudo-ratio method³⁹, assuming an *in situ* dissociation constant of 1.1 μM for Fluo-4⁴⁰. Contractions were optically recorded simultaneously with Ca²⁺ transients by illuminating the cell of interest in red light ($\lambda > 665\text{ nm}$) subsequently directed to a CCD camera (IonOptix Myocam). The cell length signals were converted to voltage via a video motion director (VED 205; Crescent Electronics) and contraction amplitudes from different myocytes were normalized by calculating the per cent change in cell length.

Determination of cell-cell coupling. In experiments to assess the interconnectivity between iCMs and CMs, the whole-cell patch-clamp method was used to introduce a gap-junction-permeable (that is, mobile) and an impermeable (that is, immobile) dye into the same cell of interest within a small group of apparently coupled isolated myocytes ($n = 5$ groups from 5 independent hearts). The mobile dye was calcein (5 mM), and the immobile dye was 1 mM dextran-conjugated Cascade Blue (MW 10,000). The immobile dye was chosen to be well separated spectrally, both from calcein and from tomato, which was used for labelling iCMs (Invitrogen). The dye pair was included in standard intracellular solution (containing 5 mM EGTA) and cytoplasmic loading was allowed to proceed for 2 min, after which the patch electrode was withdrawn from the patched myocyte. The sarcolemma of the cell resealed after pulling off the pipette, aided by high EGTA levels in the filling solution, thereby trapping the dyes in the cytoplasm. Blue fluorescence from the immobile indicator was excited at $365 \pm 40\text{ nm}$, whereas calcein fluorescence was excited at $470 \pm 40\text{ nm}$. Fluorescent images were recorded using IonOptix Myocam via a video frame grabber (#166VCB, Hauppauge) for processing using ImageJ software.

The functionality of cell-cell junctions was investigated by imaging the intercellular transmission of Ca²⁺ waves and excitation between myocytes pre-loaded with Fluo-4 AM as detailed above for the field-stimulation studies. Superfusion with 2 μM ouabain for 5–10 min was used to induce intracellular Ca²⁺ overload accompanied by Ca²⁺ wave activity, and videomicroscopy revealed the spatio-temporal relationships of Ca²⁺ waves translocating within and between the individual cells imaged in small, adherent groups ($n = 6$ groups from 5 independent hearts).

Quantitative RT-PCR. iCMs were manually sorted based on the presence of fluorescent lineage markers. Approximately 100 iCMs were pooled for RNA isolation. Similarly, ~ 100 endogenous CMs isolated using the standard Langendorff apparatus (see above) and ~ 500 cardiac fibroblasts using the migration assay (see above) were prepared for RNA isolation. RNA was extracted by the TRizol method (Invitrogen). RT-PCR was performed using the Superscript III first-strand synthesis system (Invitrogen). qPCR was performed using the ABI 7900HT (TaqMan, Applied Biosystems) as per the manufacturer's protocols. Optimized primers from the Taqman Gene Expression Array were used.

Statistical analyses. Differences between groups were examined for statistical significance using unpaired Student's *t*-test or ANOVA. A *P*-value < 0.05 was regarded as significant. Error bars indicate standard error of the mean (s.e.m.).

38. Kurrelmeyer, K. M. *et al.* Endogenous tumor necrosis factor protects the adult cardiac myocyte against ischemic-induced apoptosis in a murine model of acute myocardial infarction. *Proc. Natl Acad. Sci. USA* **97**, 5456–5461 (2000).
39. Cheng, H., Lederer, W. J. & Cannell, M. B. Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science* **262**, 740–744 (1993).
40. Ljubojević, S. *et al.* *In situ* calibration of nucleoplasmic versus cytoplasmic Ca²⁺ concentration in adult cardiomyocytes. *Biophys. J.* **100**, 2356–2366 (2011).